BEHAVIORAL AND BRAIN SCIENCES (2002) **25,** 297–388
*Printed in the United States of America*

# The self-organizing consciousness

**Pierre Perruchet** and **Annie Vinter**
*Université de Bourgogne, LEAD/CNRS, 21000 Dijon, France.*
**pierre.perruchet@u-bourgogne.fr** **annie.vinter@u-bourgogne.fr**
**http://www.u-bourgogne.fr/LEAD/**

**Abstract:** We propose that the isomorphism generally observed between the representations composing our momentary phenomenal experience and the structure of the world is the end-product of a progressive organization that emerges thanks to elementary associative processes that take our conscious representations themselves as the stuff on which they operate, a thesis that we summarize in the concept of Self-Organizing Consciousness (SOC).

**Keywords:** Associative learning; automatism; consciousness; development; implicit learning; incubation; language; mental representation; perception; phenomenal experience

The conventional cognitive framework rests on the existence of a powerful cognitive unconscious. Indeed, most psychological models heavily rely on the possibility of performing manipulations and transformations of unconscious representations using algorithms that are unable to operate while accommodating the functional constraints of conscious thought.

This target article explores the viability of an alternative framework which has its origins in the work of Dulany (1991; 1997). In this alternative, "mentalistic" framework, to borrow Dulany's terminology, the only representations people create and manipulate are those which form the momentary phenomenal experience. The main challenge is to explain why the phenomenal experience of adult people consists of perceptions and representations of the world which are generally isomorphic with the world structure, without needing recourse to a powerful cognitive unconscious. Our proposal is that this isomorphism is the end-product of a progressive organization that emerges thanks to elementary associative processes that take the conscious representations themselves as the stuff on which they operate. We summarize this thesis in the concept of self-organizing consciousness (SOC).

We first provide evidence of self-organization in the context of an experimental example that concerns the progressive extraction of words from an artificial language presented as an unsegmented speech flow (e.g., Saffran et al. 1997). Our approach is supported by a computer-implemented model, PARSER, the details of which are presented elsewhere (Perruchet & Vinter 1998b). A remarkable feature of PARSER is that the only representations generated by the model closely match the conscious representations people may have when performing the task. We then show that, provided that we accept a few simple assumptions about the properties of the world that are likely to capture subjects' attention, the rationale underlying PARSER may be extended to the discovery of the relevant units that form natural language and the physical world and also account for word-object mapping.

We then apply the same principles to more complex aspects of the world structure. We show how the SOC framework can account for some forms of behavior seemingly based on the unconscious knowledge of the syntactical structure of the surrounding environment. This demonstration, which was originally stimulated by the literature on implicit learning of arbitrary structures, finds some echoes in the literature on language processing (notably in the so-called distributional approaches, e.g., Redington et al. 1998), problem solving (for instance, in the computation/representation trade-off proposed by Clark & Thornton 1997), incubation (e.g., Mandler 1994), decision making, and automatism (notably in the instance-based models, as proposed by Logan 1988 and Tzelgov 1997a). We also show how the SOC framework, in conjunction with simple additional hypotheses, readily accounts for transfer between event patterns across sensory content, as shown for instance in the Marcus et al. (1999) study.

Finally, we argue against the empirical reliability of some additional phenomena that seemingly require the action of the cognitive unconscious. In this context, we critically examine the studies reporting that implicit memory

Pierre Perruchet is working at the Centre National de la Recherche Scientifique and is affiliated with the Université de Bourgogne (France). He received his PH.D. from the Université René-Descartes in Paris with a thesis on classical conditioning. Since then, his research has focused on automatism formation, priming, and implicit learning His main research interest is the relation between elementary forms of learning and conscious processes and representation.

Annie Vinter is Professor of Developmental Psychology at the Université de Gourgogne (France), a position she has held since 1992. She received her Ph.D. Psychology from the University of Geneva (Switzerland). Her research is directed to understanding developmental processes especially as it applies ot sensorimotor and representational development. She is currently more particularly interested in the relationships beween implicit learning processes and development.

© 2002 Cambridge University Press 0140-525X/02 $12.50

https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

and implicit learning can occur without any attentional processing of the material during the familiarization phase (e.g., Cohen et al., 1990; Eich 1984), and the data allegedly demonstrating the possibility of unconscious processing of semantic information (e.g., Dehaene et al. 1998). Issues related to the apparent dissociation between performance and consciousness in neuropsychological syndromes, such as blindsight, are also briefly discussed.

Our analysis leads to the surprising conclusion that there is no need for the concepts of unconscious representations and knowledge and, a fortiori, the notion of unconscious inferences: Conscious mental life, when considered within a dynamic perspective, could be sufficient to account for adapted behavior. This alternative framework is more parsimonious than the prevalent conceptions in cognitive and developmental sciences because it manages to account for very sophisticated behavior while respecting the important constraints inherent to the conscious/attentional system, such as limited capacity, seriality of processing, and quick forgetting (and even takes advantage of these constraints).

## 1. Questioning the cognitive unconscious postulate

In this introductory section, we point out that, in contradiction to the widespread idea that the issue of consciousness is computationally irrelevant (sect. 1.1), the prevalent computational view of mind is grounded on the postulate of an omnipotent cognitive unconscious, which has been tacitly present from the very beginnings of the information processing tradition (sect. 1.2). We then outline an alternative perspective, in which this postulate becomes useless (sect. 1.3). This section ends with the presentation of our objectives and an overview of the paper (sect. 1.4).

### 1.1. The computational view of mind

The objective of psychology, in the prevalent computational view of mind, is to study how human subjects process information. It should be noted that this objective includes no mention of the status, conscious versus unconscious, of the processed information. Addressing this issue is generally conceived of as unnecessary. Indeed, the nature of the representations and computations included in these models make that they could, *in principle,* be either conscious or unconscious when implemented in a human brain. This contention holds, irrespective of whether this processing is construed in terms of rule abstraction and application as in the mainstream tradition or in terms of multivariate statistics computation as in the connectionist approach. The fact of being conscious or unconscious is, for a mental construct, a property that does not affect the way this construct intervenes within a processing sequence.

The function assigned to consciousness, when it is considered, generally consists in making certain parts of cognitive functioning accessible. To quote 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." In keeping with this metaphor, the states and operations involved in information processing models occur in the same way whether they are concurrently accessed or not. This type of speculation is often summarized in the claim that consciousness is "computationally irrelevant."

### 1.2. The cognitive unconscious

We agree with the claim that qualifying as conscious or unconscious a representation or an operation involved in a computational model has no effect on the way this model works. But 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 most information processing models necessarily rely on a cognitive unconscious, for at least two reasons. First, the algorithms forming the models rarely match the phenomenal experience of the subjects running the tasks that, presumably, trigger these algorithms. Second, and more importantly, these algorithms are generally unable to work while accommodating the functional constraints of conscious thought, such as limited capacity, seriality, relative slowness of processing, and quick memory decay. As Lewicki et al. (1992) wrote to emphasize the power of the cognitive unconscious: "Our conscious thinking needs to rely on notes (with flowcharts or lists of if-then statements) or computers to do the same job that our nonconscious operating processing algorithms can do instantly and without external help" (Lewicki et al. 1992, p. 798).

Chomskyan psycholinguistics provide a striking illustration of these points. Whatever the fuzziness of the operational measures of consciousness, it is not tenable that the *conscious* mind is endowed with a Universal Grammar, makes assumptions about the properties of the ambient language, and tests hypotheses in order to set parameters at their appropriate values. To a lesser extent, similar remarks can be addressed to most information processing models. For instance, it is quite common to assume the existence of a syntactic processing device. Even the discovery of words in the continuous speech stream has been conceived of as the product of a mathematical algorithm of optimization, performed thanks to a statistical inference method (e.g. Brent 1996; see below, sect. 3.4). Some untaught rules of spelling are also assumed to be unconsciously abstracted (e.g., Bryant et al. 2000).

Of course, the premise of a cognitive unconscious is not limited to the studies on language. Let us consider the transcoding of numerals. One of the most influential transcoding models (McCloskey 1992) assumes that all numerical inputs are translated into an amodal and abstract representation of quantity, associating every number to a power of ten (e.g., 4030 should be coded $(4)10^3$, $(3)10^1$). Motor activities are also of concern. For instance, how do fielders modulate their speed up to catch a ball before it reaches the ground? According to McLeod and Dienes (1993), they run so that $d^2(\tan\alpha)/dt^2 = 0$, where $\alpha$ is the angle of elevation of gaze from fielder to ball. The authors wrote: "Children probably discover this somewhat obscure strategy . . . by extrapolating from their experience of watching balls thrown towards them . . . This strategy is obviously not available consciously. That its effectiveness is


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

discovered demonstrates the power of the brain's unconscious problem-solving abilities" (McLeod & Dienes 1993, p. 23).

These few examples make it clear that, by construing information processing as the main target of psychological science, regardless of the conscious status of the processed information, the prevalent view does not remain neutral with regard to the consciousness issue. This view rests in fact on the existence of a *cognitive unconscious* (Kihlstrom 1987; Shevrin & Dickman 1980). By this expression, we mean that the prevalent view takes for granted the existence of unconscious representations, together with the possibility of performing unconscious manipulations and transformations on these representations. By the same token, the concept of cognitive unconscious includes the assumption that the notions of unconscious knowledge and memory are meaningful, and most authors would probably add to this list the notions of unconscious rule abstraction, unconscious analysis, unconscious reasoning, unconscious inference, and so on.

### *1.3. An alternative framework*

**1.3.1. The "mentalistic" framework.** This paper explores the possibility of an alternative framework, in which the cognitive unconscious has no place. Mental life is posited as co-extensive with consciousness. This idea, in fact, is not new. It has even occupied a respected position in the philosophical tradition since Descartes. More recently, this framework has been cogently articulated by Dulany (1991; 1997), who called it, for want of a better term, the "mentalistic" framework.

The mentalistic view does *not* challenge overall the notions of representations, and the idea that rule abstraction, analysis, reasoning, and inferences can be performed on these representations. Conscious experience of each of us provides direct evidence for such operations. This evidence supports the conservative conclusion that we abstract rules and makes various computations and inferences when we have direct experience of doing so. These aspects of mental life, that Dulany (1997) calls the "deliberative episodes," are not of focal concern in this paper, although we do not intend to play down their importance in any way.

A departure from the standard cognitive view arises when there is no conscious evidence of performing the cognitive operations that a psychological model stipulates. As pointed out above, the lack of concurrent subjective experience is not thought of as a problem in the information processing tradition, because consciousness is thought of as providing only an optional access to the product of unconscious computations. By contrast, the mentalistic view rejects the notions of unconscious rule abstraction, computation, analysis, reasoning, and inference. Because unconscious representations have no other function than to enter into these activities, eliminating the possibility of these activities actually makes the overall notion of unconscious representation objectless.[1] Accordingly, the most salient feature of the mentalistic framework is the denial of the very notion of unconscious representations. The only representations that exist, in this view, are those that are embedded in the momentary phenomenal experience.

Representations, of course, are generated by neural processes, of which we are unaware. Thus, in the mentalistic framework, mental life comprises only two categories of events: The conscious representations and the unconscious processes generating those representations. The two are linked like the head and the tail of a coin. To quote an earlier paper of ours: "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 1999a, for a link between the notions of representation and consciousness.)

**1.3.2. About terminology.** Common sense knowledge of notions such as process, representation, and computation, even if difficult to constrain within an exhaustive definition (as is the case for many other concepts), appears sufficient at this point, because the originality of the mentalistic perspective is anything but a matter of subtle terminological nuances. However, it may be useful to *exclude* one particular understanding of the notion of representation and computation.

It has become increasingly common to define any pattern of neural activity as a representation, especially in the connectionist framework (e.g., Elman et al. 1996, p. 364). Given this approach, any biological consequence of the presentation of a stimulus is a representation of this stimulus. For example, the projection of the world on the retina of the eye provides a representation of the world. A logical consequence of this definition is that most representations are fully unconscious. Of course, such a definition has its own internal consistency: From the observer's point of view, retinal images are indeed world representations. However, the meaning of the concept is different in the mentalistic framework. Throughout the present paper, the word "representation" designates a mental event that assumes the function of some meaningful component of the represented world (e.g., a person, an object, a movement, a scene) within the representing world. At least two functions can be envisaged (Dulany 1997). A representation may evoke other representations (the representation of a pencil may evoke the representation of a pencil box, an exercise book, and so on). It may also enter as an argument into deliberative mental episodes (the representation of a pencil may be involved in reasoning, inference, action planning, and other mental activities). In this terminology, the retinal projection of the pencil does not represent the pencil, because the mosaic of cells of the retinal surface activated by the light reflected by the pencil does not fulfill any of these functions.

Likewise, the notion of computation does not extend to any neural activity, but instead designates the mental operations that take representations as arguments. In the following section, the term "computation" will be taken to be synonymous with expressions such as "computation on mental representations."

**1.3.3. An illustration.** The concrete implications of endorsing a mentalistic view will now be illustrated using a very simple situation. Let us assume that a stimulus S1, initially neutral with regard to its behavioral consequences, comes to elicit an avoidance reaction after its repeated pairing with an aversive stimulus S2. Everyone will have recognized here the schema of a classically conditioned reaction. A first interpretation may be that people have acquired some knowledge about the S1-S2 relationships, then draw the inference "If S1 then S2," thus triggering an avoidance reaction when S1 is displayed. This is a version of the expectancy



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

theory of conditioning, first proposed by Tolman (e.g., 1932), and nowadays largely accepted. This view is compatible with a mentalistic standpoint, as long as people have explicit knowledge about the S1-S2 relationships, and have explicitly drawn the inference that S2 is likely to occur when S1 occurs.

Let us now assume that people no longer remember the earlier S1-S2 pairings during the test, thus making impossible the explicit inference that S1 will be followed by S2. Experimental data suggest that a conditioned reaction can still occur in these conditions (e.g., Gruber et al. 1968). In the standard cognitive view, the loss of explicit memory does not matter. People are now assumed to rely on their *implicit* memory of the S1-S2 pairings, and make the *unconscious* inference "If S1 then S2." Such an adjustment causes no difficulty, given that the presence or the absence of consciousness is held to be computationally irrelevant.

This interpretation obviously violates the premise of the mentalistic framework. However, is this interpretation mandatory? It is worth remembering here that an alternative interpretation of conditioned performance was proposed long ago. During the training phase, some subjects' perceptual experiences comprise (at least) some features of S1 endowed with the negative valence triggered by S2. Elementary associative mechanisms are then sufficient to ensure that a negative valence becomes a new intrinsic property of S1. The conditioned avoidance reaction, in this interpretation, is directly elicited by S1. The crucial point is that the formation of knowledge about the stimulus relationships is no longer involved: the link between S1 and S2 has no need to be stored in memory and remembered, either explicitly or implicitly, nor exploited through inferential reasoning. What changes with training is the intrinsic representation of S1, which becomes negatively valenced.

There is overwhelming evidence that both interpretations are needed to account for all the reported conditioning data. Some paradigms certainly trigger one process more than the other. For instance, Garcia et al. (1977) tease apart the behavior of rats preparing to cope with a painful reinforcer signaled by an auditory stimulus (a situation mainly involving the knowledge of the S1-S2 relationships), and the behavior of rats that acquire an aversion to a flavor previously associated with sickness (a situation mainly involving a change in the intrinsic representation of S1). However, most paradigms are presumably able to generate both forms of responding. In the discussion published in the pages following Garcia et al.'s (1977) contribution, Seligman distinguishes the learning of an "if-then" relationship from the acquisition of a hedonic shift. These two processes, even though very different, are both generated, according to Seligman, by Pavlovian situations. The responses elicited by these mechanisms differ from each other on a variety of experimental variables in a consistent way (e.g., in their sensitivity to the precise timing of events, in their resistance to extinction, and so on), thus strengthening the idea that conditioned behavior has a dual nature (e.g. Holland 1980; Konorsky 1967; for an overview, see Perruchet 1984).

The dual nature of conditioned responses makes it possible to encompass all the available data within a mentalistic framework. When the knowledge of the stimulus relationships is consciously represented, conditioned responses may be of one or the other form. When explicit knowledge is no longer available, however, there is no need to invoke an unconscious analog to our conscious mode of reasoning. Responses may be due to a change in the intrinsic representation of S1. In this case, there are only successive conscious experiences, with S1, initially neutral, acquiring the negative valence initially induced by S2, through the action of unconscious associative processes. Most of the conditioning literature is consistent with this interpretation. It appears, in particular, that those conditioned responses that are endowed with characteristics typical of the responses due to the formation of knowledge about the stimulus relationships, are closely linked to the conscious knowledge of these relationships (for detailed arguments, see Perruchet 1984).[2]

In this example, it is easy to understand how the same observed behavior – a conditioned response without concurrent awareness of stimulus contingencies – can be explained either in a standard cognitive view relying on the cognitive unconscious, or alternatively in a mentalistic framework which eliminates this postulate. The subsequent sections are devoted to the objective of assessing whether very complex adaptive behavior, commonly taken as indicative of unconscious rule abstraction or other unconscious computations on cognitive representations, can also be accounted for in another way, without introducing much more than the principles set out above for the conditioning data.

**1.3.4. A hopeless project?** At first glance, the weight of the empirical evidence runs against the view presented above as the behavior under examination becomes more and more complex. The main supporting argument is that most current psychological models accounting for complex behavioral phenomena rely, with indisputable success, on the existence of unconscious representations and computations.

The fact that the models based on a cognitive unconscious work might seem to negate the potential interest of an alternative model. However, the argument is not as straightforward as it might seem. Indeed, computational algorithms are so powerful that they can simulate virtually any phenomena, without proving anything about the computational nature of the actual mechanisms underlying these phenomena. Computational algorithms generate a perfect description of the rotation of the planets around the sun, although the solar system does not compute in any way. In order to be considered as providing a model of the mechanisms actually involved, and not only a simulation of the end-product of mechanisms acting at a different level, computational models have to perform better than alternative, noncomputational explanations. The point is that the comparison needed to reach such a conclusion has never been conducted.

As asserted above, the possibility of a powerful cognitive unconscious has been embedded within the principles of the information processing tradition from its very beginning, without being clearly articulated and hence without being directly challenged. Given these conditions, the current focus on the notion of cognitive unconscious appears to be simply the consequence of making earlier tacit postulates explicit. To summarize, although the pervasiveness of the concept of a cognitive unconscious and its overall success can hardly be disputed, the demonstrative power of these arguments is undermined by a hidden circularity.

### *1.4. The objectives of the article*

It is worth pointing out from the outset that our project does *not* consist in showing that the prevalent computa-


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

tional framework is unwarranted, for any logical or empirical reasons. This objective would entail demonstrating that consciousness is necessary for any form of representation and computation. But there are major obstacles facing any such demonstration. There is no *theoretical* reason for claiming that representations and computation need to be conscious. Moreover, it is difficult to conceive of any form of *empirical* demonstration. Indeed, addressing the question of the necessity of consciousness for any mental construct requires us to demonstrate that unconscious representations and computations do not exist, and demonstrating nonexistence is beyond the reach of any empirical investigation. Our aim is to assess the viability of a mentalistic view, instead of directly questioning the prevalent framework. This leads us to address a different issue, presented below.

**1.4.1. Necessity versus sufficiency.** Let us start from a twofold consideration. On the one hand, we know that at least some mental events are conscious, because we have direct and personal evidence of their existence. Even those who argue that consciousness is epiphenomenal cannot reject this assessment (although a few philosophers have questioned the very existence of consciousness; see Rey 1991, and the refutation of Rey's position by Velmans 1991). On the other hand, the existence of an unconscious mental life is a postulate or a presupposition. This presupposition is so deeply ingrained in our modern culture that it is taken for granted by most people. But the fact remains that we have, by definition, no direct proof of an unconscious counterpart to our conscious mental life. It emerges from these two premises that the mentalistic framework is more parsimonious than the prevalent view, because it exclusively relies on the representations and the mental operations we are aware of, whereas the prevalent view postulates, in addition, a parallel cognitive apparatus.[3]

In this context, questions about consciousness, in striking contrast with the overwhelming practice, may be framed in terms of *sufficiency,* rather than *necessity.* As a consequence, the question we address is: "Is it sufficient to rely on the transient and labile representations that form one's momentary phenomenal experiences, when the conventional framework commonly assumes that a large number of representations are stored in mind and manipulated in various unconscious operations?"

The example in section 1.3.3 illustrates the point. We do not argue that subjects are unable to build and use unconscious knowledge about the S1-S2 contingencies on the grounds that consciousness should be *necessary* for these operations. What we do show is that this hypothesis is only one among several possible interpretations of the fact that conditioned reactions persist beyond the forgetting of the S1-S2 contingencies. Positing that the affective reaction elicited by the occurrence of S1 has evolved during training due to unconscious associative processes is *sufficient* to account for the data.

**1.4.2. A major objective and some additional issues.** The major part of this article, namely, sections 2 to 7, will be devoted to the presentation of a new model, called the SOC Model, with SOC standing for *self-organizing consciousness.* This expression is a short-cut, and as such, it is potentially misleading. It might suggest that we intend to address the hard issues commonly linked to the notion of consciousness, such as the problem of knowing how neural events generate conscious mental states. In fact, this article focuses more modestly on the *contents* of consciousness, such as they can be described at an informational level.[4] We propose that conscious contents are endowed with self-organizing properties, which make it possible to account for a wide range of adaptive phenomena that are commonly considered to be mediated by the cognitive unconscious. Our objective is to suggest that most of the phenomena of interest for cognitive scientists can be accounted for by this model, which avoids any recourse to the concepts of unconscious representations and computation.

The next to last section (sect. 8) will deal with somewhat different issues. For quite obvious reasons, the SOC model is not devised to account for data that we consider to lack a justifiable empirical basis. However, such data may constitute an *a priori* reason for some readers to reject our approach. Section 8 addresses such phenomena, and notably the data allegedly demonstrating the possibility of unconscious processing of semantic information. We also briefly discuss, in this section, the apparent dissociation between performance and consciousness observed in a few neuropsychological syndromes, such as blindsight.

## 2. The notion of self-organizing consciousness (SOC)

In the first section, we presented an outline of how a mentalistic framework could account for a response apparently based on unconscious memory and inference, taking as example a specific finding from the conditioning area. We now have to address a far more difficult challenge, namely, to account for the most complex aspects of behavior on which contemporary cognitive science focuses. Our approach comprises two steps. The first step consists in showing that a large number of phenomena that seemingly require unconscious rule abstraction processes, inferences, analyses, and other complex implicit operations, can be accounted for by the formation of conscious representations that are isomorphic to the world structure. The second step concerns the formation of these representations, and more precisely the causes of their isomorphism to the world structure. We suggest that this isomorphism is the end-product of a self-organizing process. The general ideas underpinning these two steps will be briefly outlined in turn in this section, then developed at length in the following sections.

### *2.1. Complex conscious representations account for seemingly rule-governed behavior*

**2.1.1. Trading representation against computation.** Complex and integrative representations, we argue, make rule knowledge objectless. 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). The above discussion concerning certain findings in Pavlovian conditioning (sect. 1.4) provides a first insight about the meaning of this claim. As shown above, the change in the intrinsic representation of S1, and notably the fact that this representation, initially neutral, becomes affectively valenced during the training phase, may replace, at a functional level, the formation of the knowledge of the S1-S2 contingency and the logical inference "if S1 then S2."



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Although often indirect, supporting evidence for a representation/computation trade-off can be found in various areas of psychology. Examples include the instance-based model of categorization (e.g., Brooks 1978), the so-called episodic (e.g., Neal & Hesketh 1997) or fragmentary (e.g., Perruchet 1994) accounts of implicit learning, the notion of mental models in problem solving (e.g., Johnson-Laird 1983), and the memory-based theory of automatism (Logan 1988). Although they evolved in at least partial independence, 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. However, they subscribe to metatheoretical assumptions that are somewhat different from those of the mentalistic framework, notably with regard to the way they handle the notions of representation and consciousness. In keeping with the mentalistic framework, we assume that the representations involved in each case are conscious.

This position, we argue, increases the a priori plausibility of the representation-based views, and expands their explanatory power, for at least two reasons. First, if the momentary phenomenal experience is the only mental event, the whole power of the neural system may be recruited for its construction. Second, the construction of a representation can profit from the presence of the momentary sensory input, instead of relying exclusively on the internal, memory capacity of the brain. The growing literature on change blindness and other related phenomena (e.g., see review in Noë et al. 2000) leads us to emphasize the importance of this factor, on the grounds that perceptual experience may be more dependent on the real world than previously thought. If, for instance, a visual scene is changed in such a way that the perception of a movement is prevented (e.g., changes occur during an eye blink, or an ocular saccade, or if a blank mask is inserted between the two displays), changes are surprisingly difficult to notice. Such phenomena indicate that the world could play the role of an "outside memory" (O'Regan 1992) in the formation of the perceptual experience, hence exempting the brain from the need to retain a detailed representation of the world. These factors make the task of constructing the representations composing the current phenomenal experience considerably easier than the task of forming the permanent and ready-to-use internal model of the world required in the prevalent view of mind.

**2.1.2. The isomorphism between the actual and the represented world.** In order to solve problems that, at first glance, require rule abstraction and complex computation, a representation has to be isomorphic to the world structure. And indeed, by and large, phenomenal experience provides an internal representation of the world that is isomorphic to its structure. We generally perceive continuous speech as a meaningful sequence of words, the visual environment as composed of persons and objects, and so on. In some sense, the adapted nature of conscious representations is not a speculative and optional proposal, but derives from the most fundamental principle of evolutionary biology: as pointed out by Velmans, "if the experienced world did not correspond reasonably well to the actual one, our survival would be threatened" (Velmans 1998, p. 51). If one adheres to the views outlined above, the structural isomorphism between our conscious representations and the world is the major phenomenon we have to explain. However, some preliminary comments are warranted to make it clear that this isomorphism is not perfect, and does not need to be so.

First, the representations we create are limited by sensory constraints. For instance, we do not have any perception about the sounds outside of the 20–20,000 Hz range, and our eyes are able to detect only a very small bandwidth of the electromagnetic spectrum from around 370 nm to around 730 nm. Likewise, phenomenal experience does not provide us with any direct representation of the structure of the physical world at other scales, such as atomic microstructure or galactic organization.

Second, even the parts of the world available to our sensory equipment may be represented only partially, or even erroneously. The fact that our representation of the surrounding world does not include the whole scene currently available to our sensory equipment, but instead is limited to a narrow focus, has been recently documented in the visual domain by the studies on change blindness alluded to above. Examples of misrepresentation are also plentiful. The sun rays at the day's end are seemingly divergent in all directions whereas they are in fact (nearly) parallel, and star constellations at night have no physical reality due to the varying distances of their elements from the earth. In addition, there are innumerable cases in which our representations are biased by our interests, motivations, and their relevance for survival. The phenomenal experience of the world may even be misadaptive, as in the case of perceptual illusions in which perceptual processes that are generally well-suited in natural situations cease doing their job reliably when faced with highly specific patterns. Such phenomena illustrate that percepts and representations are isomorphic to the world structure only in a limited way. For the sake of brevity, we continue to refer to the isomorphism between subjects' representations and world structure throughout this article, even though the very phenomenon we are attempting to account for can not be described as a simple term-to-term matching.

### *2.2. Conscious representations self-organize*

The main question we have to address at this point is: How to account for the fact that the content of the phenomenal experience is, even in a limited sense, isomorphic to the world,[5] if this content is not the product of a powerful unconscious processor manipulating unconscious representations? Our answer consists in considering consciousness within a dynamic perspective, that is to say, a perspective centered on learning principles. The key point is that each conscious experience triggers associative learning mechanisms that take the components of this experience as the "stuff" on which they operate. Thanks to this phenomenon, consciousness does not only serve an immediate adaptive function, but also participates in its own development, each conscious experience allowing us to improve the content of subsequent conscious experiences. We summarize this thesis in the proposal that phenomenal experience is self-organizing.

Psychological textbooks routinely point out that there are multiple forms of learning. But they also mention that associative learning is the most fundamental and primitive, maybe the form to which all other forms are reducible *in fine*. Because our framework is primarily motivated by the


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

search for maximal parsimony, we rely exclusively on conventional associative mechanisms in the following. Relying on associative principles – reminiscent of the old-fashioned behaviorist psychology for many – within a mentalistic framework centered on the concept of consciousness may appear anachronistic. However, the paradox is one of appearance only. Although behaviorism was grounded on associative principles, the reverse is not true: Associative principles can serve equally well in other frameworks. The mentalistic view provides a highly relevant integrative framework, for at least two reasons that will be considered in turn. First, there is a natural relation between associative learning and consciousness, mediated by the concept of attention (sect. 2.2.1). Second, the assumption that learning associates conscious contents implies that associations involve complex representations, a property that considerably improves the power of an association-based view (sect. 2.2.2).

**2.2.1. Associative learning and consciousness.** The issues of learning and consciousness are generally considered separately. As a case in point, "learning" is nearly absent from the indexes of the numerous recently published volumes on consciousness. However, reasons for considering the two issues jointly arise from the close link between learning and attention, on the one hand, and attention and consciousness on the other.

Attentional processes are sufficient for associative memory and learning to occur. This means that no superimposed operations – such as some forms of intentional orientation towards learning – are required. This phenomenon is known from the conditioning and skill learning experiments run during the behaviorist era. It has been subsequently "rediscovered" in the context of the level-of-processing framework in the seventies (e.g., Craik & Lockhart 1972), and more recently in the context of the studies on implicit learning (e.g., Whittlesea & Dorken 1993). The resulting picture is that many authors, using different terminologies, have proposed a view compatible with the claim that associative learning is an automatic process that associates all the components that are present in the attentional focus at a given point (Frensch & Miner 1994; Jimenez & Mendez 1999; Logan & Etherton 1994; Stadler 1995; Treisman & Gelade 1990; Wagner 1981). Associative learning and memory are nothing other than the by-products of attentional processing (see sect. 8.1 for a reappraisal of some contradictory evidence).

Now, there is a close relation between attention and consciousness. It must be acknowledged that the psychological literature offers a somewhat fuzzy picture of this relation. Across and even within domain and epoch, one term is often preferred to the other. But this preference lacks any clear justification. For instance, the methods devised to investigate *perception without attention* differ from the methods devised to investigate *perception without consciousness.* In the former, the stimuli are supraliminal but maintained outside the current focus of attention as a result of the task demands, whereas in the latter, attention is directed toward the target but stimulus quality is degraded. However, these terminological differences are linked more to historical contingencies than to theoretically rooted reasons. At the empirical level, it turns out that both kinds of manipulations lead to analogous findings (Merikle & Joordens 1997). A more general argument for dissociating the two concepts is that attention is selective whereas "consciousness incorporates both a central focus, and a rich polymodal periphery," to borrow the expression used by O'Brien and Opie (1999b, p. 191). This argument amounts to defining attention as the conceptually driven attentional mechanisms that are directed towards a specific source of information in response to task instructions. This view defines what Schmidt and Dark (1998) call the *intention-equals-attention view,* according to which participants' intention to attend exclusively to a target is sufficient to restrict attentional processing to this target. All proposals for a dissociation (e.g., Baars 1997; Velmans 1999) amount to such a confusion. However, the fact that the instructions ask participants to pay attention to a target does not prevent them from making quick attentional shifts toward nonattended information. Therefore, unless one endorses a highly restrictive definition of attended information as the informational content on which subjects are asked to focus, we see no reason to dissociate between attention and consciousness on the basis of their relative selectivity.

Accordingly, the fact that attention and consciousness refer to the same phenomenon does not mean that they are one and the same concept. Attention is generally located on the side of the processes, and consciousness on the side of the mental states resulting from these processes. As Pribam (1980) says: "'Consciousness' refers to states which have contents; 'attention' refers to processes which organize these contents into one or another conscious state." 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. Cowan 1995; Mandler 1975; Miller 1962; Posner & Boies 1971).

**2.2.2. Associative learning and complex representations.** At first glance, associative mechanisms appear to be underpowered for the function that we assign to them. Essential to our claim is the idea that the oft-mentioned limitations of associative learning principles are overcome whenever complex representations are conceived of as the stuff on which associative processes operate. The fact that complex representations can enter into associative links, and the high explanatory power of this mode of functioning, has been pointed out in the modern literature on conditioning and learning. The following quotation, borrowed from one of the leading theoreticians of animal learning, illustrates the point:

> Properly understood . . . associative learning theory is remarkably powerful. Of course, such a theory must reject . . . the restrictive assumption of S-R theory, which allowed associations to be formed only between a stimulus and a response, and should assume that a representation of any event, be it an external stimulus or an action, can be associated with the representation of any other event, whether another external stimulus, a reinforcer, the affective reaction elicited by the reinforcer, or an animal's own actions. Equally important, however, it must allow that *the representation of external events that can enter into such associations may be quite complex.* They need not be confined to a faithful copy of an elementary sensation such as a patch of red light; they may be representations of combinations or configurations of such elementary stimuli; they may even include information about certain relationships between elementary stimuli. But once we have allowed associative learning theory these new assumptions, we have a powerful account, capable of explaining quite complex behavior – including behavior that many have been happy to label cognitive and to attribute to processes assumed to lie beyond the scope of any



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

theory of learning. (Mackintosh 1997, pp. 883–84; emphasis is ours)

However, by and large, the fact that associative principles apply to complex representations has not been exploited, and hence the power of associative learning theory has not been fully appreciated. The symbolic framework assigns a minimal role, if any, to associative processes, and most of the connectionist models, although rooted in associative principles, only consider associations between the input units of the network, which code the material piecemeal (note that the so-called constructive methods overcome this limitation, e.g., Fahlman & Lebiere 1990).

To summarize, we propose that basic principles of associative learning and memory allow conscious representations to reach their high degree of organization and adaptiveness, provided that we consider that associations occur between the rich content of conscious experiences. The notion of self-organization excludes any organizing cognitive systems or principles that would be superimposed on phenomenal consciousness.[6] The phenomenal consciousness itself ensures its own improvement in representational power, thanks to the propensity of conscious representations to evolve in accordance with basic associative learning principles. Because consciousness is an unavoidable companion of our daily life, this means that every life episode has a learning function. There are no separate phases for learning and for performance: Each phenomenal experience contributes to improving people's ability to perceive and represent the genuine structure of the world in subsequent interactions.

### 2.3. Overview of the sections 3 to 7

Thus two main ideas are embedded in the notion of self-organizing consciousness (SOC). The first is that conscious representations that are isomorphic to the world structure, due to their ability to integrate various elements in a cohesive picture, can account for adaptive behaviors commonly attributed to rule-governed thought. The second is that ubiquitous principles of associative memory and learning are sufficient to account for the formation of these representations. The subsequent sections deal with these two aspects, although, in order to begin the demonstration at its logical starting point, we begin with the second one.

We start by demonstrating the self-organizing nature of phenomenal experience in the language domain. This domain is especially relevant to our position, because it is the domain in which the notion of the cognitive unconscious may be the most deeply rooted as a result of the Chomskyan tradition. In the next section (sect. 3), we show that the ability to extract the words forming an artificial language presented as an unsegmented speech flow may be accounted for as an autonomous change in the phenomenal experience of the materials, due to the action of elementary associative mechanisms. This interpretation has been supported by a computational model, the details of which are presented elsewhere (Perruchet & Vinter 1998b). Section 4 proposes a generalization of this model to word extraction in natural language, to the formation of objects, and to the word-object mapping issue.

Sections 5 and 6 introduce a generalization of the SOC framework to other dimensions. While sections 3 and 4 concern the formation of conscious representations of elements that are generally construed as the actual world units (words and objects), section 5 applies the same principles for more complex aspects of the world structure. We show how the formation of complex representations that are isomorphic with the world structure can account for some form of behavior seemingly based on the unconscious knowledge of the syntactical structure of the surrounding environment. Section 6 deals with the fact that human behavior may be sensitive to structural aspects of the world that transcend its surface features. This problem, reminiscent of the criticisms Chomsky levelled at the once prevalent current of behaviorism, is obviously crucial for the validity of our view. We show how the SOC framework readily accounts for transfer between event patterns cutting across their sensory content. Section 7 shows how the SOC framework may find some echoes in the literature on problem solving, incubation, decision making, automaticity, and implicit memory.

To sum up, these sections provide, we hope, a model of how organisms deprived of a powerful cognitive unconscious, can behave adaptively when faced with complex world-size situations thanks to the formation of structurally relevant conscious representations of these situations.

## 3. The case of word extraction

### 3.1. The word-extraction issue

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 potentially relevant factors. Analyses of the statistical structure of different languages have shown that a number of features are correlated with the presence of word boundaries, and could therefore be used as cues for segmenting the speech signal into words (see review in Jusczyk 1997; McDonald 1997). However, the question remains of how infants abstract the statistical regularities that they seemingly exploit. It cannot be claimed that these regularities are learned inductively from word exposure without falling into circular reasoning, with word knowledge being simultaneously the prerequisite and the consequence of knowledge of statistical regularities. In addition to the difficulties inherent in their exploitation, prosodic and phonological cues in any case 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. (1996b) used an artificial language consisting of six trisyllabic words, such as *babupu* and *bupada.* 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.

The participants in the study conducted by Saffran et al.


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

(1996b) were told before the training session began that the artificial language contained words, and they were asked to figure out where the words started and ended. The processes used in these conditions may be different from those involved in natural language acquisition. Two subsequent papers from the same laboratory (Saffran et al. 1996a; 1997) partially respond to this objection. In Saffran et al. (1997), the participants' primary task was to create an illustration using a coloring program. They were not told that the continuous series of syllables, which were presented as a sound background, consisted of a language, nor that they would be tested later in any way. In the subsequent forced choice test, participants still performed significantly better than chance (although performance is comparatively impaired in these conditions, see Ludden & Gupta 2000). A still more direct indication of the relevance of these data with regard to infants acquiring their mother tongue was provided by Saffran et al. (1996a), who reported studies carried out with eight-month-old infants. The infants were tested with the familiarization-preference procedure used by Jusczyk and Aslin (1995), in which infants controlled the exposure duration of the stimuli by their visual fixation on a light. The infants showed longer fixation (and hence listening) times for nonwords than for words, thus demonstrating that they were sensitive to word structure after a brief exposure to an artificial language. Overall, the studies conducted by Saffran and co-workers offer impressive support for the hypothesis that people are able to learn the words forming a continuous speech stream without any prosodic or phonological cues for word boundaries.

### 3.2. PARSER: The principles of the model

Our aim here is to show that word extraction can be explained by the action of elementary, associative-like processes acting on the initial conscious percepts, the result of which is to modify the conscious experience we have of the linguistic input.

What is the phenomenal experience of the listener of a new language such as the one used in the Saffran et al. experiments, at the beginning and end of training, respectively? When people are confronted with material consisting of a succession of elements, each of them matching some of their processing primitives, they segment this material into small and disjunctive parts comprising a small number of primitives. As adults, we have direct evidence of the phenomenon. For instance, when asked to read nonsense consonant strings, we read the material not on a regular rhythmic, letter-by-letter basis, but rather by chunking a few letters together. In a more experimental vein, when adults are asked to write down this kind of material, they frequently reproduce the strings as separate groups of two, three, or four letters (Servan-Schreiber & Anderson 1990). The same phenomenon presumably occurs when a listener is faced with an unknown spoken language, with the syllables or other phonological units forming the subjective processing primitives instead of the letters. Certainly, when hearing an unknown language at a normal locution rate, the processing of the material is usually not exhaustive. Rather, subjects pick up a chunk of a few syllables from time to time. But this difference does not alter the basic phenomenon of chunking. Chunking, we contend, is a ubiquitous phenomenon, due to the intrinsic constraints of attentional processing, with each chunk corresponding to one attentional focus.

This initial segmentation is assumed to depend on a large variety of factors. Some factors are linked to the participants. For instance, prior experience of another language may endow participants with different processing primitives. Also, the current state of attention and vigilance may partly determine the chunk size. Other factors are associated with the situation, such as the signal/noise ratio, the time parameters of the speech signal, and the relative perceptual saliency of the components of the signal. The mixture of these factors is very likely to mean that a listener's initial conscious experience consists of a succession of chunks which are different in length and content from the words of the language.

After extensive exposure to the language, the listener's phenomenal experience is presumably the experience each of us has of our mother tongue, that is the experience of perceiving a sequence of words. Our proposal is that the final phenomenal experience of perceiving 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.[7] 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?

The reason lies in the combined consideration of two phenomena. The first depends on the properties of the human processing system. The future of the chunk which forms a conscious episode depends on ubiquitous laws of associative learning and memory. If the same experience does not reoccur within some temporal lag, the possibility of a chunk acting as a processing primitive rapidly vanishes, as a consequence of spontaneous decay and/or 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 decay 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 relevance of this phenomenon becomes clear when viewed in relation 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 a part of a word, than if they straddle word boundaries. In consequence, the primitives that emerge from the natural selection due to forgetting 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. Technical details about PARSER are provided in Appendix A, and an on-line presentation of the model is available on the URL (http://www.u-bourgogne.fr/LEAD/people/perruchet/SOC.html). Simulations have revealed that PARSER extracts the words of the language well before exhausting the material presented to adults in the Saffran et al. (1996a) experiments, and the material presented to eight-month-old infants[8] in the Saffran et al. (1996b) experiments. These results were obtained with



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

an exhaustive chunking of the input. When a more realistic fragmentary processing of the material was simulated, performances were impaired, but remained fairly good. PARSER was able to reproduce the performance of actual subjects while processing only 3 to 5 percent (according to experiments) of the sequences presented to participants. This finding suggests that PARSER was able to simulate the results obtained under attention-disturbing conditions (Saffran et al. 1997), where inattentional gaps were presumably more frequent than under standard conditions. Finally, the good performance of PARSER was not limited to the trisyllabic words used by Saffran et al., but also extended to a language consisting of one- to five-syllable words (Perruchet & Vinter 1998b).

To summarize, we suggest that parsing 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 intraword rather than between-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). Note that our solution to the word extraction issue does not involve any new and specialized learning devices. The fact that complex material is processed as a succession of chunks each comprising a few primitives is supported by a large amount of literature (e.g., Cowan 2001). The unitization of these primitives due to their processing within the same attentional focus is one of the basic tenets of associative learning (e.g., Mackintosh 1975). Likewise, the laws of forgetting and the effects of repetition are ubiquitous phenomena. Moreover, the interdependence of processing units and incoming information (the nature of the processing primitives determines how the material is perceived and the nature of the material determines the transformation of the processing primitives, and so on, recursively) is consistent with a developmental principle initially described by Piaget's concepts of assimilation and accommodation (e.g., Piaget 1985). Most current theories of development, although they use different terminology, also rely on the constructive interplay between assimilation-like and accommodation-like processes (e.g., Case 1993; Fischer & Granott 1995; Karmiloff-Smith 1992).

### 3.3. PARSER and the issue of consciousness

The functioning of PARSER, like the functioning of any other computational model, does not depend in any way on the conscious/unconscious status we ascribe to its components. As a consequence, PARSER does not demonstrate that consciousness is *necessary* for word extraction. Its objective lies elsewhere. As set out in section 1.4.1, the aim of this paper is not to demonstrate the necessity of consciousness, but instead to assess whether conscious thought, although endowed with severe capacity limitations, is sufficient to account for performance. We pointed out that devising a model to simulate conscious states while respecting the properties of conscious thought introduces considerable *constraints.* The point we wish to emphasize here is that PARSER meets much of these constraints. Crucially, the only representations included in the model closely match the conscious representations subjects 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 closely 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. By doing so, PARSER demonstrates that the transient and labile representations composing the momentary phenomenal experiences are *sufficient* for word extraction, provided that simple and ubiquitous associative processes are allowed to operate on these representations. There is no need for unconscious representations, nor for any forms of unconscious computation on these representations.

It is worthy of note that the constraints inherent to conscious thought cannot be conceived of as limitations to the model. 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. This aspect of the model makes it specially relevant for a rational analysis of cognition, such as initiated by Anderson and Milson (1989). This approach contrasts with the common mechanistic explanation, in which the cognitive system is described as "an assortment of apparently arbitrary mechanisms, subject to equally capricious limitations, with no apparent rationale or purpose," to borrow Chater and Oaksford's (1999) characterization. The rational analysis of cognition shows how apparent limitations actually serve adaptive functions, due to the characteristic of the surrounding environment. For instance, the fact that memory decays gradually over time is viewed as adaptive, because it turns out that the probability for any memory components will be needed to deal with a subsequent situation also decays over time. In this way, the efficiency of the retrieval of information from memory parallels the probability of this information being recruited for adaptive goals. Although focusing on another function, our analysis follows the same approach: 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.[9]

### 3.4. PARSER and alternative computational models

As mentioned above, the primary objective of this paper is to highlight the internal consistency of a framework grounded on a set of premises which are strikingly different from those of the standard cognitive approach. This objective prevents a detailed and exhaustive comparison with alternative models. However, pointing out some differences may help to illustrate some specificities of the SOC framework, whose PARSER provides the instantiation in the word segmentation issue. To this end, we briefly compare PARSER with two other models of word segmenta-


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

tion, based respectively on a symbolic and a connectionist architecture. The comparison concerns only the basic principles of the models, given that empirical comparative analyses are not yet available.

One recent symbolic model of word segmentation has been developed by Brent and Cartwright (1996). 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. When units have been created by the system, they help to choose among different possible segmentations of the utterances. In addition, the choice between possible segmentations takes account of certain phonotactic constraints on the form of English words. This method has been applied with some success for parsing phonetic transcripts of child-directed speech into words.

Most of the connectionist models that 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 strikingly marked sawtooth 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 appears able to parse a continuous speech flow into words (for more recent models, see Aslin et al. 1996; Christiansen et al. 1998)

Needless to say, nothing in those models 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. The consequence is that the Brent and Cartwright model is grounded on the postulate of a powerful cognitive unconscious, even if there is no explicit mention of this postulate in their paper. By contrast, an SRN relies on mechanisms that, although lacking direct support (there is no evidence of a neural implementation of the error backpropagation algorithm underpinning SRN functioning, as acknowledged by Elman et al. 1996), are a little more realistic at the neurobiological level. However, the model's contents are even more distant from the learner's experience. Even the final state, namely, the representation of the input as a set of words, is not directly provided by the network: Words can only be inferred from the graded distribution of errors after learning is completed.

These remarks on alternative models can hardly be thought of as criticisms by themselves, given that these models were not devised to account for conscious experience. However, they illustrate the specificity of the SOC framework. PARSER, which implements the SOC framework in the word segmentation issue, accounts for the formation of word while closely mimicking the subjective experience of the learner, and without calling on other principles or mechanisms than the ubiquitous principles of associative learning and memory. By contrast, the alternative models rely on various postulates about states and operations we have no evidence of, while giving strictly no function to the representations of which we have direct and immediate evidence through conscious experience. The end result is that, in the alternative models of word segmentation considered here, thorny assumptions are made about unconscious operations while the content of phenomenal experience is left both unexplained and objectless.

## 4. Learning the word units

The achievement of PARSER in simulating experimental data on artificial, over-simplified languages supports the idea that conscious representations, far from being a phenomenal by-product of complex analytical processes, are capable of self organization. We now intend to show that our model provides a reasonable account of word extraction in natural language (sect. 4.1), and also extends to the formation of object representations and word-object mapping (sect. 4.2).

The general position taken in this section is as follows. On the one hand, natural conditions are far more complex than the experimental conditions considered so far, and this leads one to expect our model to perform worse in the latter case than in the former. In particular, it appears likely that relevant units represent a very restricted proportion of the potential units that may be initially perceived, and that the process of natural selection on which our model is based will not be sufficiently efficient. However, on the other hand, the complexity of natural conditions may paradoxically help to build the relevant units. To understand the reasons, we have to go back to the basic principles of the SOC framework, and notably to the role of attentional factors in unit formation. A new unit associates the processing primitives that are attended to simultaneously. With the simple



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

artificial languages considered so far, the primitives embedded within a single attentional focus at the beginning of training are randomly selected on the basis of their temporal contiguity, because there are no other guides to constrain chunking. However, natural conditions often provide clues, which are generally excluded in experimental conditions in order to achieve better control. These clues, we will show, guide the formation of the initial chunks by orienting people's attention and allow us to deal with the problem of the unmanageable number of possible units.

### 4.1. Word extraction in natural language

Natural language acquisition does not consist in identifying six words used again and again in a few minutes, but many thousands of words distributed over years. Are the principles underlying PARSER general enough to be easily applied to such different complexity and time scales? As we have mentioned, PARSER works thanks to the interaction between one property of the language and a few properties of the human processing system. There is no reason to believe that this interaction occurs only with the simplistic language used by Saffran and co-workers. The target property of the language, namely, that the probability of repeatedly selecting the same group of syllables by chance is higher if these syllables form intra-word rather than between-words components, is obviously shared by Saffran et al.'s artificial material and by any natural language. Likewise, the properties of the processing system on which PARSER relies are very general. For instance, one fundamental assumption of the model is that a cognitive unit is forgotten when not repeated and strengthened with repetition. This assumption may be taken for granted, irrespective of whether the process occurs in the few minutes of an experimental session or across larger time scales, in keeping with a long-standing tradition of research into the laws of memory and associative learning. In consequence, PARSER's principles seem to be relevant to natural as well as to artificial language. Briefly stated, the generality of PARSER is ensured by both the generality of the behavioral laws (e.g., only repeated units shape long-lasting representations) and the generality of the language property (the most repeated units are the words) on which it relies.

However, beyond the theoretical relevance of the principles, it is possible that the complexity of the situation may give rise to an insoluble difficulty. This could be the case if natural language really consisted of a continuous, uninterrupted speech flow. But natural language includes pauses. These provide natural cues for segmenting the speech flow from its very onset. Although the information is insufficient for full segmentation, it may be quite useful for children, given that child-directed language is characterized by very short utterances separated by clear pauses. Incorporating the information provided by the pauses into PARSER is straightforward: We simply need to constrain selection of the number of primitives perceived in one attentional focus in such a way that the content of an attentional focus does not straddle pauses. It is worth stressing that this change is not an ad hoc poorly motivated addition to the model. Indeed, this change is fully consonant with the SOC framework and, notably, with the importance of attentional factors. Pauses, in fact, partly determine the content of the attentional focus, because attention naturally gathers events in close temporal proximity. Furthermore, pauses are only one among many prosodic and phonological cues capable of orienting attention in natural language processing. Overall, although we acknowledge that the present version of PARSER is certainly underpowered to deal with natural language, the principles that it implements are general enough for us to be optimistic about achieving an improved version exploiting the multiple cues which are likely to constrain the selection of the primitives embedded in each attentional focus.

### 4.2. The representation of objects and the word-object mapping issue

PARSER was initially built to account for the segmentation of a continuous speech flow observed in the experiments by Saffran and her co-workers. Saffran et al. (1999) recently showed that both adults and eight-month-old infants succeeded equally well at segmenting nonlinguistic auditory sequences. Of course, there is no reasons to restrict the applicability of the principles underpinning PARSER to the language area, and PARSER should therefore be a priori able to simulate the Saffran et al. (1999) data. Generalizing from this example, there is no reason not to apply PARSER's principles to nonsequential material, such as objects. Our objective now is to show that the model of word extraction described above is able to account for the formation of object representations.

The idea that learning is crucial for object representation has been proposed earlier in the literature, especially by Schyns and co-workers (e.g. Schyns et al. 1998). These authors show cogently that low-level object features can change with experience, thus altering the immediate appearance of objects. These views suggest an account of object perception strikingly different from the prevalent ones. Indeed, most developmental psychologists postulate that children's ability to parse continuous sensory input into discrete objects is made possible because there are some innate constraints and certain domain-specific knowledge (Bower 1979; Karmiloff-Smith 1992), assumptions (Markman 1990), presuppositions, or intuitive theories (Spelke et al. 1992) about the structure of the world, a position that naturally follows from the standard cognitive view outlined in section 1.

Some adaptations are warranted if we are to achieve our objective. Accounting for the formation of object representations implies a change in the primitives of the system, which will no longer be the syllables or other phonological units, but, for instance, spatially oriented features. Likewise, the natural principles guiding the initial chunking of primitives will no longer be temporal proximity, but spatial contiguity. However, instantiating these adaptations confront us with a problem, which arises from the fact that the number of initial units is much greater than with linguistic material. Indeed, in the auditory speech flow, the number of possible units is limited by the sequential nature of the speech 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. By contrast, a visual display can be decomposed into a virtually unlimited set of different parts, even if each part includes only spatially contiguous elements. Under these conditions, the formation of relevant units would appear to be an intractable problem.

This problem, again, finds a solution in the idea that units


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

are formed by the concurrent attentional processing of a small number of primitives. The point is that infants' attention is 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 (Bronson 1982; Haith 1978; Vinter 1986). If several elementary features move concurrently, they have a high probability of being both attentionally processed by infants, and belonging to the same real object (of course, many objects do not move; however, it is imaginable that the perceived movement generated by eye displacement in a 3-D visual field makes it possible to generalize this phenomenon to motionless objects).

The logic applied to the segmentation of the linguistic input into words and to the segmentation of the world into objects may be extended to word/object mapping. Note that the potential problem raised by the number of candidate units is exacerbated here. In real life, infants may capture within a single attentional focus unrelated componential aspects of the environment, such as a sound frequency, together with the orientation of a segment of a visual display. To illustrate the latter issue, let us consider an example inspired by a question raised by Karmiloff-Smith (1992, p. 40). When an adult points to 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? A solution based on the selective role of attention still works. What is likely to become associated is what captures the infant's attention, that is, essentially, what is new and/or moving. Presumably, considering the auditory input first, "cat" is 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.

Of course, the process of mapping as described above may sometimes fail. The infant may be quite familiar with cats, and yet surprised by the russet color of the fur of this specific cat. We predict that, in this case, the infant would mismap the word "cat" to the color russet. 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 the infants, and more generally, what they infer to be their present object of attention. On the other hand, 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 the final achievement. Our model of learning is precisely adapted to extracting signals from noises. In general, the correct mapping will be the final outcome, because the infants will hear "cat" for animals that are not russet, and will hear "russet" for animals that are not cats.

To summarize, our model of learning, initially applied to the word extraction issue, suggests a new account of infants' basic ability to parse the physical words into objects and to map words and objects. The apparent problem posed by the unmanageable number of potential units that can be initially perceived finds a simple solution thanks to the fact that attention is naturally captured by a tightly defined set of events. Of course, this account, in its present form, is just a first draft of a more complete developmental model. Such a model should address many other points. For instance, as a rule, a word does not designate a specific object or animal, but a category of objects or animals. It is easy to imagine how the phenomenon may be encompassed in a framework based on the laws of associative learning and memory. Differences between specific instances of, say, cats, can be viewed as noise for the system, whereas the common features are located in the to-be-detected signal. When the word "cat" is associated with different instances of cats, idiosyncratic features of the animals, because they are not repeated, disappear from the representation while common features are reinforced.

## 5. From lexicon to syntax

Up to now, we have proposed an interpretation for the formation of conscious representations of parts of the world, such as words and objects. However, the existence of linguistically or physically relevant representations is not commonly considered as sufficient to account for human behavior. Representations are generally construed as the elementary bricks of thought and complex human behavior is assumed to rely on the formation of some kind of abstract knowledge, in which the bricks are combined on the basis of some organizing (e.g., logical) principles. For example, in the language domain, there is a conventional distinction between the lexicon and the syntax. Both of them are assumed to be mediated by different neural mechanisms, and the role of language exposure in the acquisition process is conceived of as very different: Although some impact of learning in word acquisition is acknowledged even by strong nativists, the acquisition of grammar is attributed to innate and specialized modules. Needless to say, we do not deny that adult humans are able to abstract rules. The very existence of sciences such as logic, physics, and linguistics, testifies to the human ability to abstract the structure of complex environments. Since this section is devoted to language, it is important to point out from the outset that we agree with the contention that humans can achieve genuine knowledge of the syntax of their language. However, in the mentalistic framework, the formation and manipulation of abstract knowledge is restricted to conscious activities.

Our proposal is that the notion of self-organizing consciousness offers a way of thinking about rule-governed behavior in cases where no conscious rule analysis is performed, without having recourse to the notion of unconscious rule abstraction. The idea is that the separation between basic units on the one hand, and rules governing those units on the other, or between lexicon and syntax in linguistic terminology, is warranted in a scientific approach (i.e., from the observer's viewpoint) but has no relevance for the processing system. The purpose of the processing system is to generate a representation of the world that integrates all the momentary input (internal and external) into



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

a coherent and meaningful scene. This complex and integrative representation, we will argue, makes rule knowledge objectless. The mentalistic framework is specially well-suited to pushing the representation/computation trade-off (Clark & Thornton 1997) to its ultimate end. Indeed, claiming that representations exist only in the momentary phenomenal experience is primarily restrictive when contrasted with the conventional cognitive approach, which postulates that innumerable representations are stored and processed in parallel in the cognitive unconscious. But there is a positive counterpart. If there is no cognitive unconscious, the full power of the neural system may be mobilized for the formation of the current phenomenal experience. This opens up the possibility of generating a multifaceted and highly complex representation of the world.

Our demonstration of the ability of conscious representations to account for improved performances in rule-governed situations starts in the context of artificial languages generated by a simple finite-state grammar (sect. 5.1). Then we turn to natural language. Section 5.2 is an attempt to generalize, on a speculative basis, the principles whose efficiency has been demonstrated in connection with finite-state grammars. Section 5.3 indicates a few directions in contemporary psycholinguistic research that also exploit the ability of lexical representations to explain apparent rule-based, syntactical abilities. We then turn away from the field of language and examine the studies in implicit learning that exploit nonlinguistic material (sect. 5.4).

### 5.1. Studies involving artificial grammars

In the artificial language considered in section 3, which was used by Saffran and co-workers (Saffran et al. 1997), the subject's task is to discover the lexicon. There are no syntactical constraints, insofar as the words of the lexicon are displayed in random order. By contrast, in the situations considered in the literature on artificial grammar learning, the discovery of the lexicon raises no particular problems because the units of the language match some subject's processing primitives. However, the combinations of these units are governed by syntactical rules, which are the to-be-learned components of the situation. In most cases, the situation involves a set of consonants, the order of which is governed by a finite-state grammar, such as that initially introduced by Miller (1958). The finite-state grammars have been extensively used by Reber (e.g., Reber 1967) and many other researchers (e.g., Dulany et al. 1984; Shanks et al. 1997) working in the implicit learning field (for reviews Cleeremans et al. 1998; Reber 1993).

In a conventional situation, participants are first exposed to a set of consonant strings following a finite state grammar such as that represented in Figure 1, without being asked to learn the rules or even being informed of the structured nature of the material. A subsequent test is performed in order to reveal whether participants have learned about the grammar. This test generally consists in asking them to judge the grammaticality of new strings. The usual outcome is that participants are able to classify the new strings as grammatical or ungrammatical with better-than-chance accuracy, whereas they lack conscious knowledge about the grammar. The initial conclusion of these studies was that mind is endowed with an unconscious information processing device able to abstract the rules governing the experimental material, and then applies these rules in other

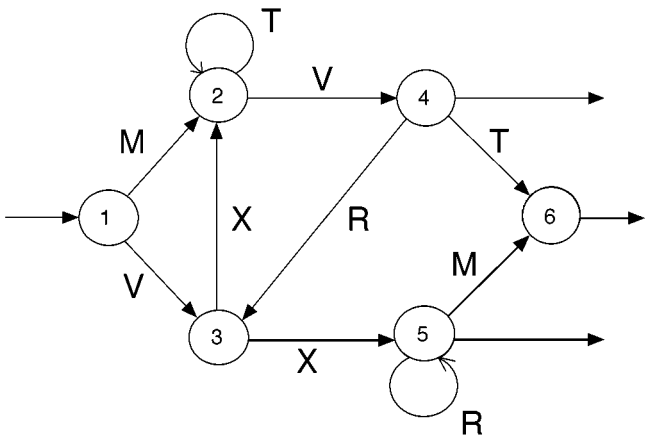



Figure 1. Schematic diagram of the grammar used in several earlier studies (Dulany et al. 1984).

contexts (Reber 1967). Because the conclusions of these early studies accorded well with the prevalent Zeitgeist, this interpretation has gone unchallenged for many years.

However, further studies, initiated by the seminal papers by Brooks (1978) and Dulany et al. (1984) made it clear that these conclusions were premature. To borrow the distinction proposed by Smith et al. (1992), the early studies failed to distinguish between a system that *follows* rules from one that simply *conforms* to rules. A ball falling on the ground conforms to the law of gravity but does not follow this law. Experimental evidence in implicit learning situations shows that the participants conform to the rules underlying the situations but there is no proof that the rules have been learned in any way. Several alternative interpretations have been proposed. Because this literature has been reviewed extensively elsewhere (e.g., Berry & Dienes 1993; see also the *Handbook of Implicit Learning* edited by Stadler & Frensch 1998), we will focus on our own interpretation.

In keeping with the SOC framework, our re-interpretation (e.g., Perruchet & Vinter 1998a; Perruchet et al. 1997b) of the phenomenon is that the training phase modifies the way the data are consciously coded and perceived. Assuming, for the sake of illustration, that XRX is a frequent recursion in the finite state grammar, participants no longer perceive X and R as two familiar, but separate entities, but perceive XRX as an increasingly familiar unit. One possible explanation for the more than chance grammaticality judgments of a new string including XRX is that participants interpret, more or less automatically, the level of perceptual fluency as an indicator of grammaticality. Strings that can be easily read because chunks of letters are directly perceived as familiar units would tend to be judged as grammatical. In short, in our re-appraisal, the formation of the conscious unit XRX replaces the unconscious extraction, retention, and use of a rule such as: If XR, then X.

It might seem, at first glance, that any fragment of a grammatical utterance is itself grammatical, and can be recombined with another fragment to form a new grammatical string. Given this logic, the initial chunking of the material would not matter. And indeed the notion of "fragmentary knowledge" conveys the tacit implication that it is a quite impoverished form of knowledge. This view is faulty, as may be illustrated using the example of natural language.


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

For instance, in the preceding sentence, "this view," or "natural language" form structurally relevant sequences, in the sense that they can be recombined with a large number of other sequences, whereas "faulty, as may" cannot be easily integrated as a component in another linguistic context, although it is a component of a legal sentence. It is obvious that it is preferable to become familiar with the former sequences than with the latter.

Likewise, in the letter strings generated by a finite-state grammar, it is preferable to become familiar with a subset of sequences – for instance, those that are generated by a recursive loop – than with other, randomly selected, sequences. We (Perruchet et al. 2002) have shown that participants in an artificial grammar learning setting indeed formed the structurally relevant units. They were asked to read each string generated by a finite state grammar and, immediately after reading, to mark with a slash bar the natural segmentation positions. The participants repeated this task after a phase of familiarization with the material, which consisted either of learning items by rote, performing a short-term matching task, or searching for rules. The same number of total units was observed before and after the training phase, thus indicating that participants did not tend to form increasingly larger units. However, the number of different units reliably decreased, whatever the task during training. This result was taken as evidence that participants' processing units become increasingly relevant as training progressed (see also Servan-Schreiber & Anderson 1990). Perruchet et al. (2002) also showed that PARSER, the computer model that was used previously to account for the discovery of words in an unsegmented speech flow (Perruchet & Vinter 1998b; see sect. 3), also accounted for participants' actual performance. Thus, the principles that make it possible to discover the lexical units of an artificial language built from the random concatenation of words, also proved to be efficient in the discovery of the syntactically relevant units of an artificial language built from a finite-state grammar.

It is worth examining why such simple principles work well in a situation that was once thought of as involving grammatical rule abstraction. It is because first-order and second-order dependency rules capture virtually all the structural constraints of the standard finite-state grammars. For instance, Perruchet and Gallego (1997) have demonstrated that consideration of only the first-order dependency rules is sufficient to account for the performance of the participants in the Reber (1976) experiments and many others which use the same material. Indeed, assuming that participants classify test items as grammatical if they consist only of permissible bigrams (whatever their location in the strings) would result in the production of 90 percent correct responses, a success level that greatly exceeds observed performance. The same demonstration may be repeated for other standard situations of implicit learning, such as the repeated sequence tasks (Perruchet & Gallego 1997).

Note that we have dealt *separately* with the lexical level (in sect. 3) and the syntactical level (in this section), while language acquisition implies the simultaneous acquisition of lexicon and syntax. This does not constitute a problem. The starting point for PARSER is the idea that each attentional chunk includes a small number of primitives, and that the primitives that are processed together form a new internal primitive, as a by-product of their joint attentional processing. After having discovered the words forming the artificial language used in the Saffran et al. (1996a; 1996b; 1997) experiments, PARSER obviously goes on creating new units. These units, which are the concatenation of a few words, rapidly vanish. Indeed, because word order is random in Saffran et al.'s material, the repetition of the same word sequence is not frequent enough to allow the strengthening of any word sequence. Let us now suppose that, instead of being randomly ordered, the words are subjected to some syntactic constraints. The constraints would make some sequences grammatical and the other sequences ungrammatical. In this case, PARSER forms long-lived units consisting of the grammatical sequences. Moreover, PARSER should discover the most frequent multi-word sequences, which have much chance of being the most syntactically relevant. If we transpose the results from the computational model to the level of the phenomenal consciousness of actual people, it appears that the same process that permitted word formation during the initial stage of learning is able to generate the phenomenal experience of well-formedness for syntactically correct word sequences. This phenomenal experience can be the source of various overt behaviors, such as grammaticality judgments or verbal productions.

### 5.2. Learning syntax in natural language

Of course, it is premature to claim that the above outline is directly relevant to natural languages. First, it may be argued that any approach relying on associative learning mechanisms can *in principle* provide only statistical approximation to genuine syntactic knowledge, whereas people make no errors. We believe that this objection amounts to both underestimating a priori the power of associative mechanisms and exaggerating the actual accuracy of people-performance. For instance, we mentioned above (sect. 3) that PARSER, although relying only on associative learning mechanisms, was able to extract the words in Saffran et al. (e.g., 1996b) language without any errors. Admittedly, this language is oversimplified but, at the same time, a very limited amount of exposure to the material is sufficient to learn it. The level of performance that can be reached when a more complex language is studied over a more extended period is currently a matter of speculation. On the other hand, people's ability to master the syntax of a natural language 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 syntactical structure of a complex sentence whenever semantics cannot help. To conclude, assessing the ultimate explanatory power of associative mechanisms is a matter for further empirical investigations and computational studies.

However, there is a second category of objections, stemming from the fact that the finite-state grammars used in the laboratory studies provide a poor analog for the grammars of natural languages. The finite-state grammars used in the implicit learning literature mainly involve first-order and second-order dependency rules between contiguous elements. By contrast, natural languages involve higher-order dependency rules and remote dependencies. At a more qualitative level, it has long been known that the grammars of natural languages cannot be conceived of in terms of a finite-state grammar. Also, it remains unclear how our claims account for other aspects of syntactic knowledge,



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

and especially the abstraction of syntactic classes such as nouns and verbs.

The part of the argument based on the consideration that our account works well only with first- and second-order dependency rules is not as problematic as it might seem. Indeed, in PARSER, the dependency rules are captured through the formation of new processing primitives, which can themselves become the components of subsequent primitives. Thanks to this possibility of hierarchical processing, we can speculate that PARSER should become at least partially sensitive to high-order dependency rules. However, the order of the dependency rules is only one aspect. Many other aspects of natural language have no counterpart in artificial languages governed by a finite-state grammar. We acknowledge that a model designed to deal with artificial languages cannot deal with natural languages without undergoing substantial changes. But the essential question is: Beyond the limitations of PARSER in its current implementation, are the fundamental principles underlying the SOC model able to account for the acquisition of syntax in natural language? Although we have no definitive response, we believe that there are arguments allowing us to answer this question in the positive.

As an example, let us consider the dependencies between remote elements, and more precisely, the case of a sequence AXB, in which A and B are associated irrespective of the length and nature of X. There are many occurrences of such a structure in natural language. For instance, in the sentence: "The window of my office is open," "the window" (A) is associated to "is open" (B) irrespective of the determinant: "of my office" (X), that may be deleted or replaced by an infinite number of subordinate propositions. PARSER is a priori unable to capture the relation, because the model posits that 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. It is in keeping with our general approach of assuming that a new unit may be composed of spatially or temporally remote events, provided that there is some reason for those events to become associated in phenomenal experience. It is easy to imagine several developmental sketches accounting for how two remote events can be joined in an unitary experience. For instance, a link between A and B may emerge in situations where both events are contiguous (a case which, in our example, corresponds to the most simple utterance: "The window is open"). Then the occurrence of A without its usual successor may result in the retention of A in a temporary buffer until B occurs in order to complete the percept AB. At this moment, A and B will be simultaneously held in the attentional focus despite their objective separation, thus providing conditions favoring both the strengthening of their association and the understanding of the sentence. This is again consonant with the SOC framework, which relies on the assumption that perception is shaped by earlier representations.

### 5.3. Converging lines of evidence from psycholinguistic research

Although they developed completely independently of our own framework, there are a number of directions in psycholinguistic research that are able to help us consider the question of language learning within the SOC framework. As an example of such work, the re-emergent distributional approaches to language have recently shown that abstract classes and categories are often associated with simple statistical properties that make them tractable by all-purpose statistical learning mechanisms. Interesting to note, even simple properties such as co-occurrence statistics turn out to be informative about syntactic classes. For instance, Redington et al. (1998) studied a large natural language corpus taken from the CHILDES database (MacWhinney 1995), comprising over 2.5 million words of adult speech. They measured the information that the context of a given word provided about the syntactic category of this word (among 12 possible categories). Context was defined by the two words to either side of the target word. The authors showed that "highly local contexts are the most informative concerning syntactic category and that the amount of information they provide is considerable" (Redington et al. 1998, p. 452; see also Gasser & Smith 1998). Distributional approaches have also proven to be able to account for other aspects of language, such as the development of word meaning (McDonald & Ramscar 2001).

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 2000b). Of course, "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 language processing necessarily involves syntactical rules. Such a view confers a high degree of probability on one of the main propositions of this article, namely, that it may be possible to explain the apparent use of abstract rules in terms of the formation of complex representations.

### 5.4. Unconscious rule processing outside of the language area

Thus far, we have focused on studies on artificial or natural languages in order to illustrate the idea that apparent rule processing may be reducible to the formation of complex representations. The same idea can be illustrated in other fields. In particular, this idea finds strong support in the literature on implicit learning that is not based on linguistic material.

Outside of the artificial grammar settings, studies on implicit learning have primarily involved two situations: the so-called serial reaction time (SRT) situations, and the control of complex systems. Most of the SRT studies have been designed on the basis of Nissen and Bullemer's (1987) par-


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

adigm. A target stimulus appears on successive trials at one of three or four possible positions, and participants are asked to react to the appearance of the target by pressing a key on the keyboard that spatially matches the location of the target. Unknown to the participants, the same sequence of trials is repeated throughout the sessions. Under these conditions, participants usually exhibit a reliable improvement in performance when compared with a control group presented with randomly generated series. The tasks involving the control of complex and interactive systems have their origin in Broadbent's studies (e.g., Broadbent 1977). Participants are placed in front of a computer simulating a complex system, such as a city transport system. Unknown to them, the parameters of the system are governed by a linear equation. The task consists of regulating the system, that is, they have to manipulate a number of parameters in order to reach and maintain a prefixed target state of the system. Several studies have shown that the initial abstractionist account of performance improvement involved unnecessary assumptions, because alternative interpretations based on simpler memory processes proved to be sufficient (see, for example, Cleeremans & McClelland 1991; Marescaux et al. 1990; Perruchet & Amorim 1992; Perruchet et al. 1997a; Shanks & St. John, 1994; Stadler 1992; Whittlesea & Dorken 1993).

Rather than examining in detail the findings resulting from these conventional situations, we focus below on a specific paradigm initially designed by Lewicki et al. (1988). Like almost all other studies in the field, this paradigm serves our primary objective which is to show that what is initially interpreted as compelling evidence of unconscious rule abstraction can also be explained in terms of the formation of conscious percepts and representation that are isomorphic with the structure of the material. However, this specific paradigm also was chosen because it allows us to illustrate another point, namely, that our interpretation can work even in cases where there is no obvious relationship between the actual rules generating the structure of the material and the participants' conscious processing units. The point is that we may be sensitive to surface regularities that are a remote by-product of the rule, so remote in fact that the logical link between the rules and their by-products may be quite difficult to discover. This subsection is dedicated to those skeptical readers who doubt the power of our approach because of their failure to understand how it can apply after a cursory examination of certain complex situations.

In the Lewicki et al. (1988) paradigm, participants were asked to perform a four-choice reaction time task, with the targets appearing in one of four quadrants on a computer screen. They were simply asked to track the targets on the numeric keypad of the computer as fast as possible. The sequence looked like a long and continuous series of randomly located targets. However, this sequence was organized on the basis of subtle, nonsalient rules. Indeed, unbeknown to participants, the sequence was divided into a succession of "logical" blocks of five trials each. In each block, the first two target locations were random, while the last three were determined by rules of the form: "If the target describes a movement $m$ while it moves from location n-2 to n-1, then it describes a movement $m'$ from location n-1 to n." Depending on whether n is the third, fourth, or fifth trial of the logical block, if $m$ is horizontal (or vertical and diagonal), $m'$ is vertical or diagonal (or horizontal or diagonal, or horizontal or vertical, respectively). It should be noted that to discover these second-order dependency rules, participants must inevitably segment the whole sequence into a succession of five-trial subsequences. That is to say, any trial within the long displayed sequence must be identified as the first, second, . . . , fifth trial within the logical five-trial block to which it belongs.

The results obtained by Lewicki et al. were clear. The participants were unable to verbalize the nature of the manipulation and, in particular, they had no explicit knowledge of the subdivision into logical blocks of five trials, which was a precondition that had to be satisfied if they were to grasp the other rules. However, performance on the final trials of each block, the locations of which were predictable from the rules, improved at a faster rate and was better overall than performance on the first, random, trials. Lewicki et al. (1988) accounted for these results by postulating that the structuring rules were discovered by a powerful, multipurpose unconscious algorithm abstractor.

Perruchet et al. (1990) provided the basis for a radically different interpretation (for an alternative interpretation based on connectionist modeling, see Cleeremans & Jimenez 1998). Perruchet et al. demonstrated that participants learned the task without ever performing the segmentation of the sequence into logical blocks. Instead, they were sensitive to the relative frequency of small units, comprising two or three successive locations. Some of the possible sequences of two or three locations were more frequent than others, because the rules determining the last three trials within each five-trial block prohibited certain transitions from occurring. In particular, an examination of the rules shows that they never generated back and forth movements (i.e., $m'$ is never the inverse movement of $m$). As a consequence, the back and forth transitions were less frequent *on the whole sequence* than the other possible movements. The crucial point is that these less frequent events, which presumably elicit longer reaction times, were exclusively located on the random trials. This stems not from an unfortunate bias in randomization, but from a logical principle: The rules determined both the relative frequency of certain events within the entire sequence and the selective occurrence of these events in specific trials. The validity of this interpretation was tested by deriving predictions concerning specific features of fine-grained performance from an abstractionist model, on the one hand, and from our alternative model, on the other. The empirical data clearly supported our re-analysis.

It should be noted that the subsequences of two or three successive locations considered by Perruchet et al. (1990) are presumably the events on which the subjects focused attentionally, and which formed their phenomenal experience of the task. Thus, adaptive performance may again be construed as a change of phenomenal experience due to the properties of this experience. Exposure to the material shapes the way it is consciously perceived and processed, and the modification of the phenomenal experience triggers the improvement in motor performance. What is new in this case, however, with regard to the situations examined above, is the fact that the link between the generating rules and the surface regularities that conscious coding can capture is far from obvious. In any case, the authors, reviewers, and the first readers of the Lewicki et al.(1988) paper were presumably all unaware of it.[10] The question of the relevance of this experimental example to real-world situations



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

is a matter for further speculation, but, at the very least, these findings strongly suggest that our account could be relevant in cases where at first glance it seems to be inappropriate.

## 6. Abstracting away from the sensory content

In the preceding section, we claimed that the changes in the way we consciously perceive and represent our environment may underlie some apparent phenomena of syntax sensitivity. In some cases, it is easy to see how a simple representation may replace genuine rule knowledge. For instance, it is easy to see how perceiving XRX as a unit may replace the rule: "If XR, then X." In the situations investigated by Lewicki et al. (1988), understanding how the same explanatory schema works is far more difficult, due to the fact that conscious processing units encode a remote by-product of the rules. But after careful scrutiny, the logic of the reappraisal is unquestionable. However, adaptation to other situations does not seem reducible to the same approach. These situations are not necessarily complex, as can be seen from the first experimental situation (Marcus et al. 1999) that we deal with below. Their common characteristics is that they reveal participants' ability to abstract away from the sensory content of the training situation, an ability that can not seemingly be explained by any association-based account.

### 6.1. Experimental evidence for abstraction

As a case in point, let us consider the recent experiments by Marcus et al. (1999). Seven-month-old infants were exposed to a simplified, artificial language during a training phase. Then they were presented with a few test items, some of which belonged to the same language while the others introduced some structural novelty. The infants controlled the exposure duration of the stimuli by their visual fixation on a light. Their discrimination was assessed through their longer fixation (and hence listening) times for items introducing structural novelty. On all these points, the paradigm was identical to that used in the studies by Saffran which are described above (e.g., Saffran et al. 1997). However, by contrast with the Saffran and co-workers studies in which the test items consisted of the syllables which formed the training sentences, Marcus and co-workers introduced a change in the sensory content of the material.

For instance, in one experiment, infants heard 16 three-word sentences such as *gatiti, linana,* or *tanana,* during the study phase. All of these sentences were constructed on the basis of an ABB grammar. The infants were then presented with 12 other three-word sentences, such as *wofefe* and *wofewo.* The crucial point is that, although all of the test items were composed of new words, only half of them were constructed from the grammar with which the infants had been familiarized. In the selected example, the grammatical item was *wofefe. Wofewo* introduces a structural novelty in that it is generated from a concurrent ABA grammar. The infants tended to listen more to the sentences generated by the ABA grammar, thus indicating their sensitivity to the structural novelty. In another experiment, infants were shown to be able to discriminate sentences generated by an AAB grammar. These results were successfully replicated in various other conditions, involving systematic counterbalancing of material and careful control of the phonetic features forming the training and the test items.

Similar studies using more complex material have been performed with eleven-month-old infants (Gomez & Gerken 1999) and with adults, most of them using the artificial grammar learning paradigm. As described above, in this paradigm, participants are first exposed to a set of letter strings generated by a finite-state grammar such as represented in Figure 1. Participants' performance is usually assessed through their judgments of the grammaticality of new strings during a subsequent test phase. In some studies, the letters forming the study items are changed in a consistent way for the test of grammaticality (e.g., C is always replaced by X, B by L, etc.). Reber (1969), and several subsequent studies (e.g., Dienes & Altmann 1997; Manza & Reber 1997; Mathews et al. 1989; Shanks et al. 1997; Whittlesea & Wright 1997) have shown that participants still outperform chance level under these conditions. The principle underlying the transfer in the so-called "changed letter procedure" has been extended to other surface changes. For instance, the training items and the test items may be, respectively, auditory items and visual items (Manza & Reber 1997), color and color names, sounds and letters (Dienes & Altmann 1997), or vice versa. Successful transfer was observed in each case. Reber claimed that these results testify to the fact that participants are able to abstract the "syntax" of the displayed material, independently of the "vocabulary."

The transfer paradigm has also been used in other contexts. For instance, Wulf and Schmidt (1997) reported experiments on implicit motor learning in continuous pursuit tracking. Unbeknown to the participants, each trial during the training sessions was divided into three segments. The target moved pseudo-randomly during two segments of each trial, whereas the other segment was the same throughout the four sessions. The test session included a transfer task in which the tracking patterns were scaled differently in amplitude or speed compared to the training sessions. The authors observed that participants selectively improved their trackings accuracy on the repeated segment, and that variations in the amplitude or the timing of the target displacement during the transfer phases had no detrimental impact on performance. Wulf and Schmidt speculated, to quote:

> If the surface structure in grammar learning is analogous to the scaled versions in terms of amplitude and overall duration in the present study, then it is tempting to suggest a parallel between the learning processes in these two domains. In both, the fundamental, or "deep," structure can apparently be learned implicitly. (Wulf & Schmidt 1997, p. 1002)

At first glance, evidence for transfer between event patterns cutting across their sensory contents cannot be accounted for by any models that rely on the statistical and distributional properties of the material, such as connectionist modeling or our own model. Indeed, the formation of an associative link between, say, *ga, ti,* and *ti,* whatever its strength, seems fundamentally unable to explain transfer to *wo, fe,* and *fe,* as observed in the Marcus et al. (1999) experiments. Accordingly, Marcus et al. concluded that infants have the capacity to represent algebra-like rules and, in addition, "have the ability to extract those rules rapidly from small amounts of input and to generalize those rules


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

to novel instances" (p. 79). Pinker (1999) echoes this conclusion, and points out that "Marcus et al.'s experiment is a reminder that humans also think in abstractions, rules, and variables" (p. 41), besides their sensitivity to simple associative learning mechanisms. Demonstrations of transfer in more complex situations have elicited similar comments. For instance, Reber (1993), talking about performance in the transfer letter paradigm in artificial grammar learning studies, claimed that "the abstractive perspective is the only model of mental representation that can deal with the existence of transfer of knowledge across stimulus domains" (Reber 1993, p. 121).

### 6.2. The outline of a reappraisal

We have no problem with the claim that the evidence of transfer reviewed above is indicative of abstraction. However, we challenge the view that abstraction is indicative of rule formation and rule use and, more generally, is indicative of high-level conceptual processing. Other authors have made the same point. Regarding artificial grammar learning studies, Brooks and Vokey (1991) must be credited for the first account of transfer that does not rely on rule abstraction. More recently, the idea that transfer does not imply rule abstraction has gained support from the possibility of accounting for transfer performance within a connectionist framework (Altman & Dienes 1999; Christiansen et al. 2000; McClelland & Plaut 1999; Seidenberg & Elman 1999). Redington and Chater (2002) have also cogently argued "that surface-independence and rule-based knowledge are orthogonal concepts." In the following, we focus the discussion on our own position, although our arguments are not incompatible with, and in some respects are similar to, those of other authors. Our claim is that transfer is a natural implication of the SOC model.

Let us return to PARSER. PARSER shows how the initial conscious percept, which is generally irrelevant to the material structure, becomes increasingly isomorphic with the structurally relevant units, thanks to the elementary principles of associative learning and memory. In section 3, we considered that the initial percept exactly matched the content of the perceived stimuli. For instance, given the auditory string *badubatibu,* we assume that participants first form the auditory units *baduba, tibu,* and so on, by chunking together the auditory primitives *ba, du, ti,* and *bu,* and this assumption was sufficient to account for the data. However, it is worth stressing that this assumption is notoriously restrictive. Indeed, the primitives that enter into the associations are internal representations that only partially match the external stimuli that trigger these representations. For instance, as a result of earlier associations, the representations of *ba, du, ti,* and *bu,* involve a written component in literate people. Thus, when a new association is built between, say, the components of the auditory percept *baduba,* the new unit is not limited to the auditory domain, but naturally extends to the area of generalization of the primitive components, and especially to the visual domain. More generally, many examples of transfer originate in the fact that conscious primitives entering into the new associations are not tied to a fixed, domain-specific format of representation, but are instead often amodal, flexible, and domain-general. Conscious knowledge is represented into a cross-system code (e.g., Fodor 1983; Karmiloff-Smith 1992), a property that ensures that any conscious content possesses a certain abstractness.

Going a step further, it may also be argued that when a few syllables are perceived within one attentional focus, the resulting conscious experience is not necessarily limited to the sum of these syllables (even considering that they are represented into a cross-system code) but instead may embed some direct perception of the overall structure. For instance, *baduti* will not be perceived as *bababa* or *baduba.* The obvious difference lies in the number of repetition of the same primitives. There is no doubt that a part of the representation of *bababa* is that it consists in the repetition of the same syllable (a pattern that we refer to as a "run" below), and that a part of the representation of *baduba* is that the same syllable is repeated with an intervening syllable (a pattern that we refer to as a "trill" below). Coding a pattern as a run or a trill entails some form of relational coding, the relation involved here being the same-different relationship. Thus our assumption is that the sensory input processed within one attentional focus may also integrate some relational information.

If we take it for granted that such abstract and relational primitives are parts of conscious representations, then there is no reason not to apply the same reasoning that we applied to more basic primitives in PARSER. Abstract primitives, if they are frequently involved in the conscious perception of a given material, can emerge from noise on the basis of a selection process analogous to the one that we showed to be responsible for the formation of sensory-based, concrete representations. As is the case for concrete representations, the extraction of regularities is facilitated by the fact that, in its turn, the initial perception determines the way further material is perceived; thus, when some abstract relations have been perceived frequently enough to become perceptual primitives, they are automatically detected in the new material whenever present. However, in this case, the end-product of the process will be the emergence of representations coding the deep structure of the situation at hand, which makes transfer to other surface features natural. To oversimplify the matter for the sake of understanding, one could say that, in the conventional account, perception provides the system with a database composed of elementary, sensory-based primitives, from which the unconscious processor abstracts the deep underpinning rules. In our account, the primitives are a little more abstract and complex. However, with these new primitive units, no further conceptual operations are needed to account for transfer.

It is worthy of note that this interpretation is viable only if the coding of the incoming information in an abstract and relational format remains simple enough to be attributed to low-level perceptual processes. Admittedly, if it turns out that the perceptual primitives needed to account for the available data are, say, nested high-level order dependency rules, it would be unrealistic to claim that these primitives are directly coded by elementary perceptual mechanisms. Thus it is important to show that the available evidence of transfer can be explained in terms of the coding of fairly simple relations. In the following section, we examine the form of abstract and relational coding needed to account for the available findings on transfer. We will show that only surprisingly simple forms of coding are required. At the same time, it is equally important to show that transfer



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

would fail if the specific constraints that our approach posits are not met. This aspect will be examined in section 6.5.

### 6.3. Perceptual primitives can be abstract and relational

To begin with the most simple case, let us consider the Manza and Reber (1997) results, showing a transfer between auditory and visual modalities in the artificial grammar learning area. These authors interpret their findings as providing support for their abstractionist, rule-based view. Although the authors do not make their interpretation more explicit, we assume that their line of reasoning could be as follows. If, for instance, subjects perceive the visual sequence XMX, they abstract the knowledge that the letter X can be repeated with a lag of one letter. When they perceive XMX again, but in the auditory modality, they may experience some familiarity with the display, because the same rule applies. This interpretation undoubtedly works well. However, the phenomenon can be easily explained without having recourse to rules. It suffices to consider that there is a direct correspondence between the visual and the auditory format of the letters X and M.

It is worth stressing the differences between the two approaches. In the former case, a rule-governed pattern needs to be extracted from the visual stimuli, before being transferred to the auditory stimuli. In the latter case, matching is direct, and independent of the structure of the material. A simple thought experiment may help to clarify the differences, and, by the same token, demonstrates the irrelevance of a rule-based account. Suppose that the material is generated randomly, instead of being generated by a finite-state grammar, and thus presents no rule-governed salient pattern. For the sake of illustration, suppose that a string such as XMT is presented. In a rule-based interpretation, transfer should not occur, because a structure can not be abstracted. Now, it is quite obvious that the prior auditory presentation of XMT increases familiarity with the visual display XMT even though there is no common salient structure (alternatively, it could be argued that in XMT all the letters are different, and that this feature is a structural characteristic. In that case, a rule-based interpretation would predict equal transfer to any letter strings in which letters are different, such as DZM, a prediction that is clearly invalid).

The same comment can be applied to other studies. For example, Dienes and Altmann (1997) observed a positive transfer between colors and the name of colors, which can also be accounted for by the natural mapping between the primitives involved in the experiment. Again, transfer would probably occur even with randomly generated stimulus sequences, thus demonstrating the irrelevance of a rule-based interpretation. However, not all studies of transfer can be explained using so simple an argument. As a case in point, the above explanation does not apply to the Marcus et al. studies in which transfer is observed between, say, *gatiti* and *wofefe,* because there is no natural mapping between *ga* and *wo,* or *ti* and *fe.*

Reinterpretation of the Marcus et al. data demands recourse to another property of conscious percepts, namely the direct coding of simple relations between the components of one percept. The relation that needs to be coded is the relation "same-different," or, in other words, the only ability that infants need to exhibit is that of coding the repetition of an event. If one postulates that infants are able to detect whether two successive stimuli are the same or not, the Marcus et al.'s results are easily explained. Indeed, as pointed out by McClelland and Plaut 1999, *gatiti, wofefe,* and more generally all the ABB items, can be coded as different-same, whereas none of the other items can be coded using the same schema. AAB items are coded as same-different; ABA items instantiate a slightly more sophisticated pattern. Note that there is no indication in the data that this pattern is actually perceived as special: Considering that ABA items do not match the pattern of the other items is sufficient to account for the data. However, it does not seem to be unrealistic to assume that a trill pattern is also directly perceived when the components of this pattern can be processed within a single attentional focus. The numerous studies (e.g., Bornstein & Krinsky 1985) showing infants' early sensitivity to symmetrical displays support this assumption.

At first glance, the demonstrations of transfer stemming from the more complex situations of artificial grammar learning in adults imply the coding of far more complex relations. We now argue that in fact, as surprising as this conclusion may be, the very same abilities that we have invoked up to now are sufficient. Indeed, although finite-state grammars embed complex relations, the coding of fairly simple patterns appears sufficient to account for improved performance in transfer situations. For instance, Whittlesea and Wright (1997, Exp. 4) reported successful transfer between letters and colors in artificial grammar learning. In the experiment, five out of the 20 training items begin with a salient alternation ("RMR"). Now, it turned out that color alternation at the beginning of a string appeared in legal test items, but never in illegal test items. It is enough to assume that participants consider the test items beginning with an alternation to be grammatical, and respond at random on the others, to simulate observed performance. If we take this interpretation for granted, then transfer is easy to account for. Indeed, although there is no natural link between, for instance, R and a red square, a natural mapping may be established between the subjective unit "RMR" and "RED/ YELLOW/ RED," or any other color alternation. Again, the observation of a positive transfer is irrelevant as to whether subjects have abstracted the complex grammar used to generate the material. It can be accounted for more parsimoniously by assuming that subjective units are at least partially represented into a relational code.

For a still more complex illustration, let us consider one of the recent studies by Shanks et al. (1997), which concluded that transfer in artificial grammar learning is mediated at least to some extent by abstract knowledge. Experiment 1 used a standard changed-letter procedure, in which the letters used during study, M, R, T, V, and X, were replaced by C, H, J, L, and N respectively for the test. Shanks et al. introduced five types of violations in their ungrammatical transfer strings. The only violation that led participants to reject the strings in a forced choice grammaticality test was illegal letter repetitions. In the original grammar, only R, T, and V could be repeated. Thus, in legal transfer items, H, J, and L could also be repeated, but C and N could not. Shanks et al. showed that participants rejected transfer items including a repetition of one of these two letters at a significant level. Such a result suggests that subjects were able to perform a quite sophisticated analysis, including at least two steps. They first have to identify the fact that M and X were never repeated in the original set, then to es-


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

tablish a correct mapping between M and C, on the one hand, and X and N on the other.

It can be shown that correct responses imply neither of these steps. Let us assume that participants have formed subjective units, each composed of a few letters. An examination of the training strings shows that these subjective units include far fewer repetitions than if letters had been selected at random. The training strings included nine repetitions, whereas we assessed (through a computational simulation) the number of repetitions expected by chance at about 22. Now, looking at the five pairs of transfer strings testing the "illegal letter repetition" feature, it appears that ungrammatical test strings always include more letter repetitions than grammatical test strings. It is enough for the participants to feel that the encoding units including a letter repetition to be unfamiliar for them to choose the grammatical item from each pair. The point is that there is strictly no need to infer what letter repetitions were legal in the study strings, or to establish a letter to letter mapping: It suffices to be sensitive to the fact that subjective units rarely include a letter repetition, whatever the nature of these letters. Transfer originates in the fact that a unit's feature such as "including a letter repetition" may be captured naturally, and not in the abstraction of the rules of the finite state grammar used to generate the letter strings (for other analyses pointing out to the primary importance of repetition structure to account for transfer in artificial grammar learning, see Gomez et al. 2000; Tunney & Altmann 1999).

With an appropriate change in terminology, we believe that the studies by Wulf and Schmidt (1997), in which successful transfer was observed on repeated patterns in a motor tracking task even though the tracking patterns were scaled differently in amplitude or speed in comparison with the training session, can be easily encompassed within the same line of reasoning. Indeed, to be brief, the analogy between a small and a large movement pattern is immediate and natural. In more formal ways, the natural correspondence between the training and the transfer patterns follows from the long-standing contention that motor behavior may be subdivided into a deep, spatial-temporal structure (Schmidt's "relative timing" of movement), and a component which is scalable in terms of amplitude and rate. There is a natural term-to-term mapping between the training and the transfer patterns because movement is not encoded in absolute spatial or temporal units, but instead as a generalizable internal schema. The spatial-temporal structure, we argue, is analogous to the representations emerging from the processing of the strings of letters in artificial grammar learning: both are, in some sense, schematic, flexible, and prone to generalization, although the dimension on which generalization occurs is unrelated to the dimension involved in the generation of the rules.

### 6.4. Is our account of transfer more parsimonious?

To recapitulate, in the conventional models, the data made available to the central processor are the individual sensory-based events. The task of finding analogies between events which differ in their surface appearance is the job of some further inferential processes. These processes belong to the domain of cognition, and more precisely, because we are not aware of them, to the realm of the sophisticated cognitive unconscious. In our alternative conception, unconscious (but elementary) processes provide a conscious representation of the sensory input that is directly framed in some abstract and relational way, as any conscious content is. With this modified input, the performance observed in transfer situations no longer needs to be explained in terms of a sophisticated unconscious processor. The ubiquitous learning and memory processes evoked in the previous sections are sufficient to explain the emergence of a reliable representation of the deep structure of the material. In sections 3 and 4, we indicated how simple principles of associative learning and memory explain the emergence of conscious representations which are increasingly isomorphic to the world structure in cases where the sensory domain remains identical. When applied to more abstract primitives, the very same principles account for the discovery of the structure of the material in cases where the sensory domain is changed. Suppose, for instance, that a grammar-generated string such as ABA is naturally perceived as a trill. If this particular pattern is not repeated, this will be quickly forgotten, and other more frequent patterns will certainly emerge. However, if a trill reoccurs frequently, even under different surface features, it will become a part of subject's representation, which in turn guides the perception of the material that is displayed subsequently. Thus, where the conventional approach makes use of complex rule inference processes which are applied to unconscious representations, we propose no operations other than those driven by the ubiquitous mechanisms that are basic to our approach.

Opponents of this position might argue that our conception simply shadows or resituates the problem instead of solving it. The argument should be that positing that ongoing sensory information is *directly* coded into an abstract and relational code, is akin to taking as premises the to-be-explained phenomenon; and presumably further consideration of this initial stage of processing would indicate that it, in fact, involves the same kind of complex machinery that most authors include under the label of cognitive unconscious. This criticism is unsound, however, because the relationships we assume to be directly coded by low-level perceptual processes are considerably simpler than the abstract rules of the mainstream tradition. They are limited to a few aspects, including the same/different distinction, the properties of symmetry, repetition, and alternation and relationships along some perceptual dimensions such as *smaller than* or *brighter than.* It is not biologically implausible to assume that these relationships are coded at earlier stages of neural processing, although there is as yet no direct evidence (one exception is the direct coding of the relation *brighter than,* that is at least partially coded at the retinal level by lateral interaction between concurrent stimulations).

In the absence of more extensive neuropsychological arguments, our hypothesis finds some support in the primacy of relational coding in phylogenetic evolution. It has long been shown that animals such as rats are able to perform tasks involving elementary forms of relational learning successfully. For instance, if rats are trained with two stimuli differing in brightness in such a way that the choice of the brighter is rewarded and the choice of the darker not rewarded, they subsequently choose the brighter of two new stimuli even though the absolute brightness of the new rewarded stimulus may be identical to that of the old unrewarded stimuli. Thus, rats appear to be sensitive to the relationship between stimuli rather than to their absolute



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

properties. Such a demonstration has been replicated with various animal species and using a variety of simple relationships, such as *larger than.* Primates and a number of birds also appear able to learn a discriminative response to pairs of stimuli depending on whether they are identical or different, and once acquired, this ability transfers to any new stimulus pair irrespective of its nature.[11] Within the perspective of evolutionary biology, these results are not at all surprising. In many cases, the raw information provided by an isolated event is only partially relevant. For instance, the retinal size of a perceived object or animal is uninformative, because it depends on the distance between the observer and the distal stimulus. Similarly, the absolute brightness provides incomplete information, because perceived brightness depends on the ambient luminance. Considerably more reliable information is provided by a relational coding, by means of which the size or brightness of a new stimulus is assessed by comparison with contextual stimuli.

### 6.5. Analyzing transfer limitations and failure

A major advantage of a parsimonious account lies, somewhat paradoxically, in its limited power, which makes it easier to falsify. Indeed, our account is certainly unable to explain all possible kinds of transfer, and demonstration that these types of transfer actually occur should be taken as a compelling refutation. This section is devoted to show that transfer is in fact severely limited, as our account anticipates.

**6.5.1. The transfer decrement phenomenon.** In experiments where positive evidence of transfer is reported, performance levels on the transfer situations are, as a rule, lower than performance levels on the original training situation. This so-called *transfer decrement* phenomenon raises a problem for a rule-based standpoint. In an authoritative discussion on the use of abstract rules, Smith et al. (1992) posit as the first of their eight criteria for rule use that "Performance on rule-governed items is as accurate with unfamiliar as with familiar material" (Smith et al. 1992, p. 7; see also Anderson 1994, p. 35; Shanks 1995, Ch. 5). In the context of artificial grammar learning studies, Whittlesea and Dorken posit that "a subject who learned a useful rule would have equal success in transfer on stimuli presented either in the same or different features, because the rule is applicable regardless of the features in which items are presented" (Whittlesea & Dorken 1997, p. 66). Manza and Reber (1997) acknowledge this implication of their own abstractionist view. Thus, an essential prediction of any system that uses algebraic rules to represent its knowledge about some domain is that its transfer performance on novel items should be just as good as its performance on familiar items. The question arises: Why is the phenomenon of transfer decrement ubiquitous?

A simple way to reconcile the empirical evidence with the assumption that knowledge is rule-based is to assume that the rules are not absolute, but probabilistic, and that they have limited scope. Although this argument is logically sound, it is clear that it severely undermines the core advantage of rule-based approaches, namely, that they provide general and abstract descriptions of the stimuli. Rules that only apply to familiar cases obviously have only limited interest. In short, rules have a potential adaptive value insofar as they can be applied to novel situations. This is indeed what made them so attractive to early cognitivists such as Chomsky. Another possible explanation of the transfer decrement phenomenon in a rule-based framework would be that the usual training conditions provide insufficient practice. This explanation accords with the Manza and Reber (1997) view. According to these authors, performance is initially sensitive to low-level surface features, then becomes increasingly independent of those features, and exclusively determined by the deep structure of the material. After sensory-based representations have been built, to quote, "an 'abstractor' would come into play, gradually removing irrelevant surface elements and leaving only structural elements in the representation" (Manza & Reber 1997, p. 101). The transfer decrement phenomenon would correspond to an intermediate stage of training in which performances would reflect a mix of influences from specific and abstract components, in which the top level of the abstractive process has not yet been attained.

Pacton et al. (2001) tested this hypothesis. They reasoned that training in laboratory settings is necessarily restricted, both in duration and in the number of stimuli experienced by participants. To overcome this limitation, they tracked the time course of transfer performance over the extended durations typical of the acquisition of complex skills in natural settings. They examined the development of children's sensitivity to certain orthographic regularities based on experience of printed language. For instance, some experiments exploited the fact that, in French, the consonants that can be doubled are only doubled in the medial position of words (i.e., never at the beginning or at the end). This rule is never taught and the situation therefore taps implicit learning processes. Children became increasingly sensitive to the legal position of double consonants from grade one to grade five. However, the major point of interest concerned whether this sensitivity transfers to consonants that are never doubled in French. Rule-based approaches would predict that children learned the rule that consonants are only doubled in medial position from a subset of consonants that are seen in doublets, then transfer this knowledge to consonants that are never seen in doublets. This should result in a progressive convergence of performance, with training, on seen and unseen material.

Pacton et al.'s results clearly invalidate this prediction. There was no trend towards a reduction of transfer decrement amplitude over the five years of training that were examined. The performance curves for seen and unseen material remained parallel throughout practice. This parallelism was observed in several experiments and also applied to other orthographic rules. Overall, these results suggest that even after exposure to, presumably, several million words in which a rule applies, children's orthographic behavior still can not be readily qualified as rule-directed.

Note that the persistence of transfer decrement across extended practice is fully consistent with our view. To be fair, the persistence of transfer decrement is consistent with any view that relies on statistical or distributional properties of the material. Indeed, in such views, transfer is construed as generalization, with generalization gradients depending on the similarity between familiar and novel forms. In contrast with the predictions issuing from an abstractionist view, there is no obvious reason to expect that the amount of generalization depends on the level of training. Distributional approaches would predict continued lower levels of


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

performance on novel material, even after extensive training, because the similarity between familiar and novel situations remains the same across time. In keeping with this observation, any statistical approach is able to account for parallelism between performance on familiar and novel material over practice.

**6.5.2. Accounting for transfer failure.** Up to now, we have dealt with the results showing evidence for transfer, even if the transfer decrement phenomenon makes this evidence less powerful than abstractionist theorists would presumably hope. This emphasis on positive results is warranted. However, it is worth stressing that totally negative results are certainly the most frequent outcome in the relevant literature. With the notable exception of between-letter transfer in artificial grammar learning, transfer failure has frequently been reported in the literature on implicit learning. Total failure to obtain transfer to new material with dissimilar surface features is the rule in studies involving serial reaction time tasks (e.g., Stadler 1989; Willingham et al. 1989) or control process tasks (e.g., Berry & Broadbent 1988; Squire & Frambach 1990).[12] In the conclusion to their review of transfer in the most current implicit learning paradigms, Berry and Dienes (1993, p. 180) pointed out that "the knowledge underlying performance on numerous tasks . . . often fails to transfer to different tasks involving conceptually irrelevant perceptual changes." This empirical finding leads the authors to propose that limited transfer to related tasks is one of the few key features of performance in implicit learning tasks. Likewise, a surprising specificity of learning has been observed in the coordination between perception and action during infancy (e.g., Adolph 2000). In the literature on problem solving, which will be discussed in the next section, there is also overwhelming evidence for the difficulty of transferring the solution of a problem to another, when both problems have the same deep structure but different surface features (e.g., Clement 1994).

A model positing a powerful unconscious rule abstractor is obviously equipped to account for positive results, but, as an inevitable consequence, is undermined by negative results. Demonstrations of the empirical influence of problem content on performance have challenged the prevalent models of problem solving in the last decade, which have typically had recourse to formal or abstract rules, with a striking separation (based on the computer analogy) between rule-based programs and stored representations (e.g., Braine 1978; Cheng & Holyoak 1985). Most rule-based accounts have difficulty in predicting when and how transfer occurs and when and how transfer fails. By contrast, the SOC framework makes predictions about the conditions that are likely to promote, or hamper, the possibility of transfer. Briefly, transfer is expected only when the commonality between the training and the new situation is a part of the conscious representations triggered by the two situations. In other words, transfer is only possible when the elements common to the original and the new situation are components of the conscious percepts. More precisely, the SOC framework anticipates that transfer occurs only when subjects' attention has been focused on the common abstract features. Many results lend support to this prediction. As Reeves and Weisberg (1994) concluded in their review,

> in almost all cases, subjects must either work at schema induction by comparing the similarity between base analogues (Catrambone & Holyoak 1985; Reeves & Weisberg 1990), mapping one analog onto another (Ross & Kennedy 1990), or being explicitly provided with schematic principles that accompany the base analogues (Fong et al. 1986; Gick & Holyoak 1983). (p. 390; see also Clement 1994).

Needless to say, we are not arguing that the SOC model is the only one capable of accounting for these data (see e.g., Singley & Anderson 1989). Rather, our claim is that the findings evidencing transfer limitations and failures are compatible with this model, while they are difficult to reconcile with the idea of a cognitive unconscious giving automatic access to the deep structure of a problem.

## 7. Problem solving, decision making, and automaticity

The experimental studies presented above suggest that the formation of conscious representations which are consonant with the structure of the material, accounts for at least some of the phenomena usually attributed to processes that would operate through the sequential, analytical manipulation of information. Our aim now is to show that this suggestion may find echoes in the literature on problem solving, decision making, and automaticity.

### *7.1. Problem solving and incubation*

Each of us has direct evidence of the sequential manipulation of symbols according to certain logical rules. Indeed, the solution to a problem is sometimes obtained through the effortful elaboration of a chain of reasoning. However, in many cases, the solution to a problem springs to mind without the phenomenal experience of engaging in logic-analytic operations. Conclusions simply rise to consciousness, without being the outcome of a worked-out inference. This dual nature of reasoning was acknowledged long ago, and framed into different terminology (Sloman 1996). For instance, Smolenski (1988) distinguished between a rule interpreter and an intuitive processor. Likewise, Shastri and Ajjanagadde (1993) opposed reflective reasoning, which requires conscious deliberation, and reflexive reasoning, in which inferences appear as a reflex response of our cognitive apparatus. Johnson-Laird (1983, p. 127) talks about explicit and implicit inferences, and Hinton (1990) distinguished between complex (rational) and simple (intuitive) inferences to refer to the same distinction. We are concerned here with only the second aspect of these dichotomies, the one which taps what Dulany (1997) calls the evocative mental episodes.

**7.1.1. Problem solving as the formation of new subjective units.** In keeping with the dominant Zeitgeist, solving complex problems without the apparent involvement of explicit deliberative processes, 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 that are



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

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). The Shastri and Ajjanagadde (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 encoded as such can be directly "read," instead of being computed through analytical inference processes.

**7.1.2. Incubation.** A marginal aspect in the literature on problem solving concerns the phenomenon of incubation. Everyone has had the experience of the solution to a problem suddenly occuring after we have given up our deliberative and unsuccessful search for it. The phenomenon may happen 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." This phenomenon is somewhat difficult to investigate in the laboratory, but there is nevertheless some experimental evidence for it. For instance, Fulgosi and Guilford (1968) asked their participants to anticipate the consequences of various improbable events, either for a period of four minutes, or during two sessions of two minutes separated by unrelated activities. Delays of at least 20 minutes were beneficial in producing responses. 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 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 of response 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. For the sake of illustration, let us suppose that the correct segmentation of *batubidutaba* is *batubi/dutaba,* but that a subject initially perceives *batu/bidu/taba.* These percepts shape new internal units, and because perception is guided in turn by earlier processing units, the same display has a chance of eliciting the same erroneous perception in subsequent trials. Fortunately, internal units are progressively forgotten during the delay intervening between two repetitions. This makes it possible for a new parsing – which may turn out to be correct – to occur in subsequent trials. In one sense, one could say that, in PARSER, correct segmentation is the product of an incubation effect. Obviously, the subjective experience of mind popping is lacking with an artificial language, because a solution, whatever it is, never corresponds to a meaningful perception, as may be the case with other materials. But it is easy to imagine how the model could account for mind popping in a situation where a correct solution could be immediately identified as such, instead of being gradually confirmed with training.

To conclude this discussion of problem solving, 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. The phenomenon of incubation, which gives us strong intuitive feeling that some unconscious genius goes on to work inside our minds alongside our conscious occupations,


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

might be nothing other that the forgetting of structurally irrelevant solutions. This conclusion fits well with the conclusion reached in earlier sections about other forms of learning. It could also be expanded to other forms of learning that space limitation prevents us to 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, p. 45).

### 7.2. Decision making

Going a step further in our speculation, decision making might prove to be another area of application of our framework. Of course, as in the case of reasoning, we do not refer here to the decisions that are the products of a deliberate, step-by-step conscious analysis, but to the decisions that emerge immediately, before any rational considerations. 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 far 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 (see also the quotation of Mackintosh 1997, in sect. 2.2.2 of the present article). Thus, the conscious representations that have developed under natural conditions are probably endowed with an emotive dimension which results from self-organization and which may be directly responsible for the decision.

### 7.3. Automaticity

The terms automatic and unconscious are often used interchangeably in everyday language. This is also the case in the writings of several psychologists, such as Jacoby ( Jacoby et al. 1993). At the same time, there is a consensus on the idea that an automatic mode of responding is not limited to only the most simple situations. Combining these two premises leads us to infer the existence of complex and sophisticated unconscious processing, a conclusion that is at odds with our general framework. Which of the two premises turns out to be questionable? We have no problem with the claim that people are able to deal with complex situations in automatic ways. Reading is often designated as the archetypical example of automatism, and, irrespective of the fuzziness inherent in the concept of complexity, it must be acknowledged that acceding to the meaning of a word from its graphemic representation is anything but a simple task. However, we strongly disagree with the collapsing of the notions of automaticity and unconsciousness. To make the point clear, we need to return to the literature on automatism formation.

There is general consensus that automaticity can be defined in terms of three main criteria (e.g., Neumann 1984). The first refers to a mode of operation: An automatic process is not subject to interference from attended activities, and does not interfere with such activities. This criterion is often operationalized by the lack of interference in dual task experiments, in which participants have to carry out two actions simultaneously. The search tasks, in which participants are assumed to perform operations in parallel on a single visual display are also used for the same purpose. The second criterion refers to a mode of control: An automatic process can be triggered without a supporting intention (strategies, expectancies, etc.), and, once started, can not be stopped intentionally. The Stroop task is the preferred way of investigating this property. In the prototypical version of this task, of which many variants exist, subjects are asked to name the color of a word while ignoring the word. The time taken to identify the color when it is paired with an incongruent color word is usually found to be slower than when it is paired with a neutral word, an effect revealing that the irrelevant word has been processed without intent. Finally, the concept of automaticity is defined by a mode of representation: Automatic processes are often unconscious. All of these properties are conceived of as a consequence of extended training. A given processing, initially susceptible to interference and under subjects' conscious control progressively loses these properties during practice with the task. This general pattern of changes, which can be observed in many situations of our everyday lives, leaves us with the idea that the very same operations that are initially performed consciously come to be performed, after appropriate training, by a powerful unconscious processor operating in parallel.

To begin with, it is worth emphasizing that the above description provides an idealized view of the phenomenon. The whole literature on automatism is characterized by a few initial papers which have posited a set of definitory criteria (e.g., Hasher & Zacks 1979; Shiffrin & Schneider 1977), followed by an overwhelming number of experiments demonstrating that these criteria are never fulfilled, even in those activities, such as reading, that everyone believes to be as prototypical of automatisms. A convincing argument for the graded nature of automatisms was presented in two well-documented reviews as early as the middle eighties (Kahneman & Treisman 1984; Neumann 1984). Subsequent research has confirmed and strengthened this standpoint. Maybe we should place special emphasis on the Stroop effect, because this effect is recurrently described as a compelling demonstration that reading lies outside of people's intentional control. In a recent experimental paper entitled: "The Stroop effect and the myth of automaticity," Besner et al. (1997) report that the Stroop effect is eliminated when a single letter instead of the whole word is colored. From this and other related findings, they conclude that empirical data "are inconsis-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

tent with the widespread view reiterated in over 500 journal articles, chapters, and textbooks that a Stroop effect occurs because unconscious/automatic processes cannot be prevented from computing the semantics of the irrelevant word" (Besner et al. 1997, p. 224; see also Besner 2001; Dishon-Berkovits & Algom 2000).

However, the fact that the properties of automatisms are gradual rather than all-or-none is not sufficient to rule out the view that, as an effect of repeated practice, cognitive operations and representations progressively relax their initial link with conscious awareness. Consciousness, in this view, appears to be an optional quality of cognitive activities, a proposal that contradicts 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 Samuels (1974) theory, in which automatization is equated to 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. These theories are obviously consistent with the prevalent Zeitgeist, and 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 were challenged by Logan and his collaborators (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. This theory accounts nicely for the empirical data in which the notion of automatism is rooted. Indeed, retrieval from memory requires a minimal amount of cognitive effort and attention, and hence interferes minimally with other operations. Also, retrieval from memory is often triggered by the surrounding stimuli without any possibility of intentional control. Lastly, the nature of the process engaged to retrieve the solution is unavailable to consciousness. For instance, in the face of the problem: 5 + 3 = ?, an adult subject produces the response "8" with minimal cost, has difficulty in preventing the occurrence of this solution in mind, and has no introspective knowledge of the way by which the solution pops into his mind. These characteristics strikingly differ from those of the operations undertaken by a child performing the same addition on her fingers.

Several, although not all, aspects of Logan's theory of automaticity are directly compatible with the SOC model. To go a step further, Logan so-called "instance theory" is based on three main assumptions. The first is the obligatory encoding assumption, which asserts that attentional processing of an event causes it to be encoded in memory. The second assumption is the obligatory retrieval assumption, which asserts that attentional processing of an event causes the retrieval of whatever was associated with this event in the past. The SOC model shares the same two assumptions. However, it strongly departs from the instance theory on the third assumption. Logan assumes that each event is represented separately in memory, even if it is identical to a previously experienced event. As extensively described above, the SOC model is rooted in an associative theory of learning and memory, which provides, we believe, a far better account of the progressive tuning to the world structure of subjective percepts and representations. However, our point here is not to discuss Logan's theory further, but rather to borrow the elements of this theory that allow the SOC model to encompass the data related to automaticity.

In the SOC model, automaticity may be construed as the possibility for a subject of forming a new conscious representation the components of which were previously perceived as independent primitives. Note that this definition does not differ from the one we proposed for implicit learning. When people create a new unit such as *bupada,* this unit is also composed of initially independent primitives such as *bupa* and *da.* The difference lies in the fact that, for instance, the final unit *bupada* 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" needs 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 "*bupada.*" As Logan contends, automatic behavior is nothing other than memory retrieval.

The difference between this interpretation and the various interpretations framed in terms of attention withdrawal (e.g., Laberge & Samuels 1974) or parallel processing (e.g., Shiffrin & Schneider 1977) is overwhelming. This difference does *not* primarily refer to a simple/complex dimension. Presumably, the biological mechanisms involved in creating a single conscious perception and representation are incredibly complex (but all the resources of the neural circuitry can be recruited for this task, given that it is the only one to be performed at a given moment). The point is that these mechanisms are grounded on associative principles, and do not involve the manipulation of unconscious symbol-like representations. A particularly clever empirical demonstration that automatic behavior does not consist in performing unconsciously the very same set of operations initially performed under attentional control has been provided by Zbrodoff (1999) in the context of arithmetic problems. Zbrodoff reasoned that if skilled people pass through intermediate counts while they solve a simple addition problem (e.g., gone through 5 and 6 when they solve 4 + 3) as children do when they begin to do arithmetic, intermediate counts should have a priming effect on subsequent tasks that involve those intermediate counts. She tracked this effect while subjects practiced alphabet arithmetic problems (e.g., B + 4 = F), and found that the priming effect of intermediate counts, observable in novices, disappeared after extensive practice.

Up to now, we have dealt only with cognitive automatisms, such as reading and arithmetic calculation. In the psychological literature, as well as in everyday language, the notion of automatism also embraces the motor components of behavior. Is it possible to encompass these aspects within


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

the view outlined here? We believe that the response is yes, provided we accept the possibility that action and its results may be admitted as components of the phenomenal experience, in the same way as sensory input. This would lead to the formation of rich representations including not only our body and the world, but the interaction between them. Again, the entire literature on conditioning, and especially instrumental or operant conditioning, provides striking demonstrations that organisms' responses can enter into associative links. In keeping with our general framework, it follows that our own action and its consequences can participate in the self-organization of conscious representations, thus providing structurally isomorphic representations of the world – including ourselves and the consequences of our own actions. To take a simple example, switching the light switch while entering a familiar dark room may become a constitutive component of the phenomenal experience of this life episode.

We mentioned above that automaticity and the absence of consciousness are frequently referred to as identical (e.g., Jacoby et al. 1993). Our conclusion is, ironically, at the exact opposite. The phenomenal experience is, to a large extent, the product of an automatization process. Tzelgov recently entitled one book chapter: "Automatic but conscious: That is how we act most of the time" (Tzelgov 1997a, p. 217). We fully agree with this claim, which, unsurprisingly, Tzelgov infers from his endorsement of Logan's theory of automaticity on the one hand, and Dulany's mentalistic framework on the other. 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 the 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.

## 8. Revisiting other purported evidence for the cognitive unconscious

The primary objective of all the prior sections was to demonstrate the viability of a framework involving exclusively conscious representations and computation, by showing how the concept of self-organizing consciousness could account for a wide range of adaptive phenomena usually considered to be mediated by the cognitive unconscious. However, for obvious reasons, we have not commented on arguments that lend support to the cognitive unconscious based on data we consider to lack a reasonable empirical basis. We now have to consider these arguments. First, we will deal with the idea that one or several events could influence the processing of a subsequent event without the initial episode(s) having been attentionally processed. Second, we will turn towards the phenomenon of unconscious semantic priming, which constitutes one of the most immediate objections to be raised when the possibility is suggested, in formal setting or informal discussions, that a cognitive unconscious might have no actual existence. Finally, we will examine the literature on rare neuropsychological syndromes, such as blindsight, which also lend apparent support to a cognitive unconscious. We do not intend to provide an exhaustive discussion on these issues. Rather, our aim is to outline the way the arguments relying on these phenomena can be discounted through the detailed examination of a few examples, while referring to other discussions in the literature when available (for other critical examinations of the literature, see, e.g., Dulany, 1991; 1997; O'Brien & Opie 1999a; 1999b).

### *8.1. Implicit memory and learning without attentional encoding*

Research into implicit memory (or repetition priming) provides overwhelming evidence that processing stimuli may induce changes in performance on the subsequent identification or production of the same stimuli, without it being necessary to retrieve the initial encoding episode explicitly. For instance, when the initial event is the reading of verbal items, subsequent facilitations in the processing of these items has been reported for word completion, tachistoscopic identification, identification in a perceptual clarification procedure, and many other perceptual tasks that do not require the explicit retrieval of the initial event. The implicit memory tasks may also rely more on the encoded meaning of concepts than on the perceptual record of the items. For instance, in the category-exemplar generation test, participants are asked to name the first exemplar of a given semantic category that comes to mind. Exemplars that were previously displayed are evoked more frequently than unseen exemplars. Several studies suggest that the effect of the initial event may be observed even in cases where the explicit retrieval of this event is not only absent from the task demand, but made impossible, due either to specific experimental manipulations (e.g., a large study test interval) or to amnesic disorders due to neurological lesions (for reviews, see Roediger & McDermott 1993; for a skeptical standpoint about the experimental demonstration of the phenomenon in normal subjects, see Butler & Berry 2001).

Although they have been the object of a considerable amount of interest over the two past decades, the basic phenomena highlighted in implicit memory research are anything but new. Indeed, nearly a century of research into conventional situations of learning and conditioning teaches us that prior experiences influence behavior in subsequent situations, without subjects explicitly remembering the events involved in the original experiences. Clear-cut evidence may also be found in everyday life. To take a simple example: Each of us is able to complete 5 + 3 with the solution 8, and this ability has obviously been acquired through experience. But it is unlikely that anyone is able to evoke the original training experience. The fact that repetition priming is generally studied after a single exposure differentiates this paradigm from standard learning studies, which generally involve multiple trials, and this feature may explain why the phenomenon has been compared to the memory tasks such as recall and recognition, rather than being integrated in the field of learning. But this is merely



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

a statistical difference: Many studies investigate "one-trial learning," and other studies investigate the effect of multiple repetitions on priming. These studies confirm the conclusion that might reasonably be anticipated, namely, that the same processes are involved in both cases (see also Logan 1990).

Overall, these phenomena are fully compatible with our framework. Moreover, they provide elementary examples of the progressive transformation of conscious experiences after earlier identical or similar processing episodes, and are therefore central to our approach. It should be noted that we have assumed that this transformation was mediated by the unconscious tuning of processing mechanisms as a by-product of their recruitment, and not by the retrieval of earlier episodes. Therefore, the observation of an effect caused by a past event that is currently forgotten, whether the phenomenon occurs in normal subjects or amnesic patients (for a review, see Gabrielli 1998), does not undermine our view in any way.

However, one aspect that is potentially difficult to reconcile with our framework has been reported. Indeed, some results suggest that an earlier event can have an effect even though this event has not been attentionally processed. In implicit memory research, the study material is usually presented in fairly standard conditions, that is, without any attempt to prevent subjects from paying attention to the displayed items. Some studies have investigated implicit memory after subjects had been faced with a secondary task during the study phase and these studies have returned positive results (e.g., Parkin & Russo 1990). But these conditions were not intended to entirely prevent attentional processing. Eich (1984), on the other hand, has reported implicit memory for verbal information that was claimed to be totally ignored in a selective listening procedure. If such a result turns out to be robust, it argues against one of our basic postulates, because it suggests that the meaning of a word can be accessed unconsciously. Also, it argues in favor of a dissociation between the conscious/attentional system and learning, whereas the linkage between the two notions is a fundamental principle of the concept of Self-Organizing Consciousness.

An examination of the literature, however, leads us to doubt the reliability of implicit memory without attentional encoding during the study phase. For instance, Eich's conclusion has been challenged by Wood et al. (1997), who showed that Eich's positive results were due to the slow rate of presentation used in this study. This allowed participants to pay at least some amount of attention to the to-be-ignored channel. There is now overwhelming evidence that attention to the material at the time of encoding is a necessary condition for the observation of an effect of these materials in subsequent implicit memory tests, such as word completion and perceptual identification tasks (e.g., Crabb & Dark 1999), reading tasks (MacDonald & MacLeod 1998), or object decision tasks (Ganor-Stern et al. 1998).

The same conclusion emerges from the implicit learning area, in which interest focuses on the effect of more complex and structured situations than those involved in implicit memory research. In most studies, the to-be-learned material is displayed in normal conditions and the need for the attentional processing of this material has been acknowledged ever since Reber's early papers (e.g., 1967) on artificial grammar learning. However, the hypothesis that implicit learning could occur without attentional encoding has been proposed in different contexts. Berry and Broadbent (1988), for instance, have introduced the concept of unselective (i.e., without attention) learning. Unselective learning was assumed to occur when the situation was too complex to be solved by attention-based mechanisms. Cohen et al. (1990) also assumed nonattentional learning, although their proposal was diametrically opposed to Berry and Broadbent's position. Their hypothesis was that attention is required for learning complex sequences, while nonattentional learning is effective for the simplest forms of sequential dependencies. In both cases, supporting evidence was provided by studies that used a concurrent secondary task during the training session.

Recent studies strongly challenge the claim that two forms of learning can be distinguished, with a nonattentional form emerging when the situation is very complex (e.g., Berry & Broadbent 1988) or very simple (Cohen et al. 1990). In some cases, the prior evidence has not been replicated. For instance, Green and Shanks (1993) failed to replicate some of the results obtained by Broadbent and coworkers despite extensive attempts to do so, and observed that, as a rule, the secondary task impaired performance irrespective of the complexity of the task. In other cases, the prior evidence has been reinterpreted. Subsequent reappraisal has shown that, as in the field of memory research, the dual task conditions routinely used in the studies investigating the role of attention in implicit learning paradigms did not prevent brief attentional shifts towards the relevant information. As a case in point, Jimenez and Mendez (1999) conclude from their recent experimental studies that selective attention to the predictive dimensions in a sequence learning paradigm is necessary to learn about the sequential relationships (see also Jiang & Chun 2001, for similar evidence from another paradigm, Frensch et al. 1994; Hsiao & Reber 1998; Shanks & Channon, in press, for other approaches that emphasize the role of attention in implicit learning). Finally, the idea that any form of learning requires attentional processing receives a powerful support from studies in classical conditioning in humans, which have consistently failed to demonstrate that conditioning was independent from conscious cognition (for review, see Lovibond & Shanks 2002).

### *8.2. The unconscious processing of semantic information*

We have now to examine results suggesting the possibility of semantic representations without concurrent conscious experience. Of special relevance is the so-called unconscious semantic priming effect. The semantic priming effect designates the influence of a prime on the processing of an immediately following target, when this influence logically implies the access to the meaning of the prime, beyond its low-level perceptual features. For instance, a word prime may shorten the naming time of a semantically related target or influence the liking judgment of the target (e.g., Greenwald et al. 1996), even though the prime and the target are different. The unconscious semantic priming (USP) phenomenon corresponds to the case where the influential prime is unconsciously identified. Admittedly, the USP phenomenon requires that the system generates and uses a symbolic representation without any conscious counterpart, a requisite that is obviously at odds with the principles underpinning a mentalistic framework.


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

**8.2.1. Does the USP phenomenon have a solid empirical basis?** First, it should be mentioned that there are many reports of failures. For instance, Bar and Biederman (1998) asked their subjects to name a familiar object presented in subliminal conditions, as proven by the fact that performance was at chance in an immediately subsequent forced-choice test of recognition. This brief exposure resulted in a substantial increase in the naming accuracy of the same object later in the session. However, this effect was limited to the case where the same object was presented in the two occurrences. No facilitation was observed when the second object shared the same name but not the same shape (e.g., an office swivel chair and a four-legged kitchen chair, or a motorboat and a sailboat). A considerable decline in the effect was observed when the same object was slightly translated in the visual field. Similar failures to reveal unconscious semantic processing have been reported using binocular rivalry as a tool (for a review, see Blake 1998). When different images are shown to the left and right eyes, the conscious percept is characterized by alternating periods of left-eye dominance and right-eye dominance. The question is: How is information processed that is normally visible but suppressed from conscious awareness while the other eye is dominant? Many experiments have shown that some aspects of information processing are unimpaired, such as certain visual aftereffects. However, once again, all semantic processing is completely disrupted. Thus, words erased from consciousness by rivalry suppression failed to improve performance in a subsequent lexical decision task (Zimba & Blake 1983).

However, some authors claim they have obtained positive evidence for USP. Positive evidence has typically been obtained in conditions where the target follows the prime by a fraction of a second. For instance, in Greenwald et al. (1996), the influence of a prime on the liking judgment of a semantically related target was obtained only when the prime target interval did not exceed 100 msec. To the best of our knowledge, the fact that the effect of semantic, that is, deep encoding may be extremely short-lived, has never been considered to be an objection even though it runs counter to the most established findings in the memory field. Whatever the case, even tagged with this astonishing characteristic, the existence of USP, if confirmed, rules out our claim that the only cognitive representations are those that form the phenomenal experience.

Our argument is that a compelling demonstration of USP has not yet been provided. All the alleged demonstrations of the phenomenon have been followed by devastating criticism from skeptics. A hallmark of this recurrent sketch is the *BBS* target article by Holender (1986), who concluded from an impressive re-analysis of the available data that

> none of these studies has included the requisite controls to ensure that semantic activation was not accompanied by conscious identification of the stimulus at the time of presentation . . . On the basis of the current evidence, it is most likely that these stimuli were indeed consciously identified. (Holender 1986, p. 1)

Since then, new papers for and against the argument have been published. For instance, Draine and Greenwald (1998) recently presented a new methodology to demonstrate USP, a methodology that Merikle and Reingold, in a subsequent comment, found "compromised by the same issues concerning the measurement of awareness that have plagued all previous attempts to use the dissociation paradigm to demonstrate unconscious perception in the complete absence of conscious perception (Merickle & Reingold 1998, p. 304; see also, Miller 2000).

**8.2.2. An illustration.** The presentation of the criticisms made by Holender (1986), Merickle and Reingold (1988), and a few others, goes well beyond the scope of this article. However, to illustrate, a detailed analysis of a new example may be useful. The recent claimed experimental evidence for USP provided by Dehaene et al. (1998) is of special interest, because its recency and its publication in a high-impact journal suggest that it provides an existence proof of USP that addresses all the earlier criticisms. Dehaene and co-workers used a task in which participants had to press one key if a target number was larger than five, and another key if the target was smaller than five. The target was immediately preceded by a masked prime number, which could be also either larger or smaller than five. Thus, by crossing the values of the primes and the values of the targets, the experiment comprised four conditions, with half of them being congruent and half incongruent with regard to the expected motor response. The authors observed that congruent conditions elicited a reaction time 24 msec shorter than noncongruent conditions. Interesting to note, this positive priming effect was obtained even when (1) the numerical prime was presented as an Arabic digit and the target as a spelled-out number (or vice versa) and (2) when the trials with repeated displays were removed from the analysis. This suggested that participants were not influenced by the surface similarity between the prime and the target, but instead performed a comparison between the numerical prime and five, a task that undoubtedly taps the semantic level. On the other hand, because the authors obtained a nonsignificant discrimination performance for the prime, as measured by $d'$, they argued that the effect was unconscious. At first glance, the study does indeed support the authors' conclusion that "a large amount of cerebral processing, including perception, semantic categorization and task execution, can be performed in the absence of consciousness" (Dehaene et al. 1998, p. 599).

Unfortunately, further scrutiny of the paper leads to far less clear-cut conclusions. On the one hand, there are serious reasons to doubt that the task tapped the semantic level. Indeed, the whole study involved in fact only four numbers, two numbers lower than five (one and four) and two numbers larger than five (six and nine). Moreover, for each target number in a specific format (e.g., four, or NINE), the key-pressing task was repeated over 64 trials. In these conditions, after a few training trials, it appears quite unlikely that participants actually performed a comparison with five when the numbers were displayed, whether as prime or as target. Studies on automatism (see sect. 7.3) strongly suggest that participants quickly shifted from an algorithmic mode to a direct memory retrieval, linking one and four to the left key and six and nine to the right key (or vice versa depending on the group). In other words, after minimal experience with the task, it is likely that participants were no longer performing a comparison task, and instead proceeded with the numbers as they would have proceeded with meaningless visual patterns or sounds, that is to say, without any semantic involvement. (See Damian 2001, for an experimental support for this hypothesis with words as material.)

On the other hand, the claim for unconsciousness is also questionable. The masked prime was displayed for 43 msec.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Such a duration may be sufficient for identifying a stimulus such as a word. The hypothesis that the prime could have been consciously detected is strengthened by the fact that (1) the choice was limited to four, highly familiar primes; (2) the primes were short. Half of them were one-character long (the number in their digit format) and the others (the number written as words) comprised only a few letters; (3) each prime was repeated 64 times (in a specific format). This is of major importance. Indeed, it has been shown that the simple repetition of a subliminal stimulus in the same conditions of exposure greatly improves detection. For instance, Bar and Biederman (1998) report that the rate of correct identification of a familiar object presented for an average of 47 msec (range: 42 to 56 msec) increases from 13.5 percent in the first presentation to 34.5 percent in the second 15 minutes later. Finally (4), the stimuli serving as prime were perceived as targets under normal exposure conditions throughout the experiment. Although the study was intended to capture the effect of "prime" on "target," we should also consider the possibility that, despite the labels given to the stimuli by the experimenters, the participants may also have become sensitive to the priming effect that the target had on a subsequent prime.

Although these conditions make the nondetection of the prime highly unlikely a priori, the authors report a non significant $d'$ in two additional experiments devised to assess the rate of detection (Exp. 1) or discrimination (Exp. 2) of the prime. However, each prime was only presented 12 times (instead of 64 in the main experiment). Still more damaging to the authors' conclusion is the fact that there was a clear descriptive trend for an effect, at an even shorter duration of presentation than the 43 msec used in the main experiment. For instance, a prime presented during 29 msec. elicited 10.4 percent of hits versus 7.3 percent of false alarms. Despite this contradictory evidence, the authors concluded in favor of unconsciousness, relying on statistical nonsignificance. It is worth adding that the well-known shortcomings of conclusions drawn from nonsignificant results are specially relevant here, given the small number of observations on which the tests were based (these additional experiments were run with a smaller number of subjects than the main experiment: N = 6 and 7 respectively, instead of 12).

**8.2.3. Concluding comments.** To conclude, the Dehaene et al. study is no more conclusive than the many earlier attempts that have flourished in the literature for two decades or so. Far from demonstrating unconscious semantic priming, this study describes an effect that is neither unconscious nor semantic.

Parenthetically, our reappraisal highlights the fact that even recent alleged evidence for unconscious semantic priming may contain a substantial number of conceptual and methodological flaws, the nature of which has been pointed out many years ago. This phenomenon reveals the depth of the commitment of most investigators to the Zeitgeist, and illustrates how the prevalent view may reinforce itself circularly. For the concerns of this article, the Dehaene et al. study, when considered in conjunction with earlier studies, does not require us to reject the conclusion that there is to date no compelling evidence for unconscious semantic priming. The only effects for which there is a solid empirical basis concern, on the one hand, unconscious priming attributable to the processing of some surface property of the prime and, on the other, semantic priming associated with the conscious processing of the prime. None of these phenomena require the postulate that a symbolic representation can be created, stored, and used outside of the subject's phenomenal experience.

We focused above on the phenomenon of unconscious semantic priming, because it has been the most widely used argument in favor of the unconscious perception and representations of words. However, certain other studies have made use of the Stroop effect. For instance, Marcel (1983) and Cheesman and Merikle (1986) reported a Stroop effect without color word detection, thus suggesting unconscious access to the meaning of the color word. Unfortunately, such an effect has also been criticized (Dulany 1997; Holender 1986), and has been found to be very difficult to replicate. Thus, Tzelgov et al. (1997) showed that when the color word is displayed near the threshold, the stroop effect is observed only in trials in which participants correctly identified the word, and for participants who identified the words above chance level.

Accordingly, the question of whether the meaning of a nonidentified word can influence behavior is still open. (See the current debates between Brown et al. 2001 and Neely & Kahan 2001, and between Damian 2001 and Abrams et al. 2002). However, those who are tempted to see arguments for a cognitive unconscious in this literature should be aware of at least two points. First, isolated positive reports cannot be considered demonstrative. This is because inference is probabilistic; experiments aimed at demonstrating unconscious semantic priming are presumably widespread, and a few of them *ought* to report statistically significant effects. To be reliable, a demonstration must define a set of specific conditions in which the effect is reproducible. No one can argue that this condition is currently fulfilled. Second, available experiments provide at least one firm conclusion: If an effect does occur, it is weak and short-lived. To rely on such marginal phenomena to support the idea of a cognitive unconscious is hardly serious. The concept of cognitive unconscious is overly costly, and if it turns out that its only effect lies in barely detectable phenomena the adaptive function of which is questionable, all the evolutionary biology principles need to be reconsidered! As pointed out by Dulany (1999): "If claims for the power of a cognitive unconscious were correct, the experimental effects would be too strong and replicable for these literatures even to be controversial. No one can claim that." To conclude, we believe that available studies on the processing of unconscious semantic information fail to constitute a challenge to our framework.

### *8.3. Blindsight and other neuropsychological disorders*

In some circumstances, brain damaged people may give adapted responses to stimuli of which they deny any conscious perceptual experience. Although several syndromes fall under this head, we focus here on one of the most intensively investigated, namely blindsight. According to the standard description, patients with lesions in the primary visual cortex are able to respond appropriately to stimuli presented in regions of the visual field formerly represented by the lesioned cortex, without this responding being accompanied by a conscious visual experience. Among the preserved abilities are, for instance, the ability to detect the localization of a light point, the presence of motion, and even


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

the direction of this motion (e.g., Weiskrantz 1997). This pattern of findings runs counter to our framework, because it suggests that visual representations can be processed unconsciously, as indicated by patients' preserved performance on some visual tasks. Blindsight performance lends favor to the postulate grounding of the prevalent cognitive approach, namely, that consciousness provides an optional access to phenomena that occurs in the cognitive unconscious.

Our proposal is that the available data, however, do not provide such clear-cut evidence of dissociation as the above description suggests (for a critical examination of the blindsight literature, see Campion et al. 1983). On the one hand, it is worth stressing that performance involving the damaged cortical region of blindsight patients remains deeply impaired. To borrow Marcel's (1986) famous example, no one has ever seen a thirsty patient grasping a glass of water placed in the visual field formerly represented by the lesioned cortex. Even in the limited sample of tasks in which above chance performance has been reported, blindsight performance does not equal that of normals. This is also true of other, related syndromes. In a conclusion to a tutorial review covering blindsight, prosopagnosia, neglect, and alexia, Farah (1994) noted:

> Among all of the syndromes, there is none for which visual perception, in its totality, has been convincingly demonstrated to be normal or near normal. Therefore, there is no reason to view these syndromes as consisting of normal perception with conscious awareness merely stripped away . . . There is currently no evidence for a dedicated conscious awareness system, distinct from the systems that perform specific perceptual or cognitive functions. (Farah 1994, p. 72)

On the other hand, claiming that patients lack conscious experience is also an overstatement. Blindsight patients often report some feelings that guide their response. Surprisingly, this feeling is sometimes described as nonspecifically visual in nature, a fact that may have prompted observers to overemphasize the dissociation. But having abnormal subjective experience does not equal having no subjective experience at all. Overall, the neuropsychological disorders mentioned above certainly provide intriguing findings. However, the question of whether they illustrate more than above-chance, but degraded performance, accompanied by distorted, but still present phenomenal awareness, remains a point of debate.

## 9. Conclusion

### 9.1. Summary

The prevalent computational view of mind necessarily relies on the postulate of a sophisticated cognitive unconscious. Indeed, most psychological models require the existence of unconscious representations and the possibility of performing various unconscious operations on these representations, such as rule abstraction, analysis, reasoning, or inference. This article has explored the possibility of an alternative framework, originally proposed by Dulany (1991; 1997), which avoids any presupposition concerning an unconscious counterpart to our conscious mental life. In this so-called mentalistic framework, all mental life consists of nonconscious operations on conscious states, each of them doing nothing on its own independent of the other.

The challenge therefore consists in accounting for the relative isomorphism between the content of our phenomenal experience and the world structure, without calling for unconscious operations on unconscious representational contents. Our solution is based on the progressive transformation of phenomenal experiences as a result of self-organizing processes. Thanks to ubiquitous properties of the words of the language, and more generally of the objects of the environment, basic associative principles, when they are allowed to operate on successive conscious contents, appear sufficient to shape perceptual units that match words and objects. Moreover, with more extensive interaction with the world structure, the same processes turn out to be able to generate highly complex representations isomorphic with the world structure. These representations are themselves able to fulfill the function generally assigned to unconscious rule-governed thinking, as suggested in various recent research domains. Thus, when learning processes are given their full place, what seemed to be straightforward evidence for a cognitive unconscious turns out to be explicable through the self-organizing properties of conscious representations. In addition, we have shown that a wide range of phenomena generally thought to provide direct support for the existence of a cognitive unconscious, such as subliminal semantic activation, and various alleged examples of implicit/explicit dissociations, can also be encompassed within a mentalistic view.

### 9.2. Looking towards the future

Although much of the literature has been covered, our treatment of many issues has been somewhat cursory. We make no claim to have been exhaustive. It is possible that we may unintentionally have ignored some robust and replicable phenomena that may represent a challenge to a mentalistic account. It is incumbent primarily upon the opponents of this framework to identify these phenomena as empirical counterarguments. Pending further challenges, our provisional conclusion is that the self-organizing properties of conscious contents offer a way of triggering a far-reaching reappraisal of mainstream cognitive psychology. If this conclusion is accepted, a promising direction of research will be to further explore the SOC model and its implications in various domains, in order to consolidate and expand the scope of the mentalistic framework.

One issue of special importance relates to human development. Although a few indications of the implications of our view with regard to child development have been provided in section 4.2 (for a more extensive treatment, see Perruchet & Vinter 1998a), it is obvious that further work is needed. It is crucial for the SOC model to assess whether its underlying principles can be generalized across the entire life span, and notably at a very early stage of infant development. If the SOC model is able to succeed in this task, this model could provide the basis for a new developmental model that challenges most of the current theories. The theories that rely most extensively on nativism are obviously those that are the most concerned. But such a model would also be in sharp contrast to constructivist approaches, such as Karmiloff-Smith's (e.g., 1992) model, and the recent ideas inspired by connectionist modeling (e.g., Elman et al. 1996), notably concerning the status and the role of consciousness. Research aimed at exploring these aspects is currently in progress in our lab.

Another issue that deserves consideration concerns the



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

neurobiological implementations of the SOC model. At a first glance, our model, because it is based on the formation of associations, should be well-suited for translation into biological mechanisms. But this simplicity may be a matter of appearance only. Indeed, it is worth recalling once more that associations bear on complex representations, which form the components of the phenomenal experience. Although associative mechanisms are fairly well understood in cases where simple stimuli are involved, associations involving the complex content of conscious experiences would appear to require other explanatory schemas. One relevant approach may be the neural interactionism proposed by Roger Sperry (e.g., Sperry 1965; 1991), in which consciousness, conceived of as an emergent property of brain functioning, continually feeds back into the system from which it has emerged, thus resulting in the principle of downward causation. This principle now receives additional support from the field of dynamical system theory (see e.g., Thompson & Varela 2001).

Besides these issues, we have not commented on many other potentially relevant aspects, such as the implications of our view for neuropsychological disorders, or for clinical and applied psychology. Our hope is that our proposals will appear suggestive enough to first-hand workers in those different areas to motivate them to investigate these aspects.

## NOTES

**1.** It could be argued that unconscious representations also fulfill a simple function of storage. Each of us has the strong intuition that if the mental picture of a pencil (or any other of the countless objects or events not currently on view) can be evoked at any moment, the representation of the pencil is stored somewhere in the brain independently of its conscious instantiation. This reasoning is questionable, however. By way of illustration, let us consider the physical picture of a pencil on a slide. Everybody would agree that the picture fulfills its representative function for a human perceiver when it is projected on screen. But what about this picture when it is not displayed? It is only a pattern of colored pixels on a film. Also, interestingly, the picture may be stored in a format that does not preserve its analogical relation with the original scene, such as in the series of binary digits obtained after compression of the digitized picture. The storage format does not matter, because a stored picture has no other function than making possible subsequent generations of the picture. What is kept over time is the possibility of generating the picture again through appropriate procedures, material, or decoding mechanisms, which are not embedded in the stored picture. A more biologically relevant analogy can be found in the way the information needed to synthesize proteins is coded in the genes. The proteins are not stored in a ready-to-use format: What is stored in the RNA are the assembly instructions to generate a specific protein when the appropriate signals and conditions are present.

These illustrations make it clear that the possibility of generating the conscious representation of a past experience does not mean in any way that this representation has enduring existence as such (i.e., serves its function) in a putative unconscious system outside of its conscious and momentary instantiation. The same reasoning obviously holds for any form of knowledge, whether episodic or semantic.

**2.** There is at least one finding inconsistent with this hypothesis. Baeyens et al. (1992) observed conditioned reactions in subjects who were unaware of the S1-S2 relationships in an evaluative conditioning procedure. The point is that these responses were presumably due to the knowledge of these relationships, and not to a change in the intrinsic properties of the conditioned stimulus, because they were affected by a post-conditioning revaluation of the unconditioned stimulus. Detailed methodological considerations are necessary here to suggest how this contradiction can be resolved (see Shanks & St. John 1994, for a critical analysis of the Baeyens et al.'s results).

**3.** For most cognitive scientists, introducing consciousness into psychological modeling seemingly *increases* complexity. Taking for granted that unconscious mental activities are the basic stuff of the mind, the fact that some proportion of our cognitive activities appears to be conscious adds unwanted complexity, and confronts us with many difficult problems. Consciousness appears as the piece left over when the jigsaw has been completed. However, it is worth emphasizing that this line of reasoning holds only within the metatheoretical framework that we challenge.

**4.** The traditional collapsing of consciousness and language in many areas of research makes it necessary to emphasize that the contents of phenomenal experiences cannot be identified as verbalizable knowledge. As defined by Baars (1995, p. 6):

> The content of consciousness includes the immediate perceptual world; inner speech and visual imagery; the fleeting present and its fast-fading traces in immediate memory; bodily feelings like pleasure, pain, and excitements; surges of emotional feelings; autobiographical memories; clear and immediate intentions, expectations, and actions; explicit beliefs about oneself and the world; and concepts that are abstract but focal.

**5.** Provocative as this proposal sounds, we believe that taking the contents of phenomenal consciousness as a research target represents only a minimal departure, if any, from the current practice of experimental researchers. Indeed, a quick survey of the literature shows that the dependent variables used in most of the laboratory experiments are a direct reflection of the participants' phenomenal consciousness. Let us consider, for instance, the variables used in two fields that cannot be suspected of overemphasizing consciousness, namely, the fields of implicit memory and implicit learning. A typical request made to participants in implicit memory tasks is to say the first word that comes to mind. Likewise, a typical task in implicit learning research is to assess whether a string of letters sounds consistent with an artificial grammar. Obviously, such requests capture aspects of the phenomenal experience of the participants. Along the same lines, studies on perception heavily rely on what people report they see, hear, or feel.

**6.** In keeping with a standard usage of the notion of self-organization, this terminology is not intended to mean that the structure of the world plays no role in the increasing representativeness of conscious experiences. We thank Don Dulany for warning us about this misinterpretation. Claiming that phenomenal experience is endowed with self-organizing properties means that the properties of conscious perceptions and representations, *when considered jointly with the properties of the external world,* are sufficient to account for the growing consistency of these perceptions and representations with the world structure. What is excluded from the causal sketch in the notion of self-organizing consciousness is not the environmental structure, but the cognitive unconscious.

**7.** For the sake of simplicity, we make a rigid distinction between the notion of primitives (a set of elements that are perceived as a whole by the system) and the notion of chunks (the momentary content of the phenomenal experience, which may include one or more primitives). It should be clear that this distinction may depend on contextual features. For instance, a word, in the usual linguistic behavior of adult readers, may be considered as the relevant primitive; however, the primitives may also be the letters when the task consists of checking the spelling of a word.

**8.** Applying PARSER's principles to infants rests on the premise that consciousness is not limited to humans endowed with language. Obviously, infants' conscious experience is presumably different from that of adult humans. But the fact that the phenomenal experience differs between individuals, because it is built throughout one's life, is precisely one of the key tenets of the present article.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

**9.** Taking this speculation one step further, it may be noted that our model is also consistent with an evolutionary approach. Certainly it is unrealistic to assume that the processing constraints linked to consciousness evolved from their function in learning language because, presumably, the emergence of consciousness did not follow the arrival of language in the philogenetic chronology. However, as we discuss in the next section, the explanatory sketch outlined above for the words of the language also applies to the other units (objects, animals, etc.) of the real world. It is possible that the specific properties of conscious thought emerged by natural selection thanks to their efficiency in dealing with these natural units. The properties of languages could be due to the fact that language has evolved in such a way that it could be learned easily, given the human abilities, as pointed out by Newport (1990).

**10.** The very same argument could also be illustrated in the literature dealing with the implicit learning of invariant features. In McGeorge and Burton (1990), the task was to perform arithmetic computations on strings of four numbers. Unbeknown to subjects, every string contained the number three. In a subsequent test phase, the subjects had to indicate which of two displayed strings had previously been presented. In fact, both strings of each pair were new, but one of them contained the number three while the other did not. Subjects preferentially selected the strings containing the number three, although they were not conscious of this regularity. This finding was originally interpreted as evidence of unconscious rule abstraction. In fact, Wright and Burton (1995; see also Stadler et al. 2000) have shown that the presence of a fixed number within a string substantially decreased the probability of this string including a repeated digit. Different experimental findings have provided evidence that the subjects' performance was in fact due to their sensitivity to this remote by-product of the experimenter's rule.

**11.** The demonstration of the existence, in various species of animal, of abilities similar to those we consider to be responsible for human performance in complex transfer tasks suggests that our account successfully avoids any recourse to genuine abstraction. But the debate is not over, because it may be argued that animals also perform rule-abstraction. Talking about the animal conditioning literature, and explicitly borrowing their claim to Spence (1937), Wills and Mackintosh (1999) observed that

> relational theories are not all alike. In one version, they seem to be talking of a conceptual process that abstract relationships such as taller than, brighter than, same as, different from, and so forth. In another, they are appealing to a much lower-level sensory process that allows the contrast between the two neighboring stimuli to enhance the perceived difference between them. (Wills & Mackintosh, 1999, p. 48)

These authors favor the simpler sort of mechanisms on the basis of the limitations inherent to the phenomenon of relational coding. The argument is as follows. If relational learning is the product of a sophisticated process of abstraction operating on the basic properties of the stimuli, it looks reasonable to anticipate that abstraction will occur irrespective of the nature of these stimuli, provided they are successfully encoded. By contrast, if relational coding is the end-result of low-level, hardwired mechanisms, there is no reason to expect generality over modalities and features. In support of the latter account, Wills and Mackintosh showed that pigeons have no difficulties in learning relations between rectangles differing in brightness, whereas they fail to learn the relations between stars with different number of vertices, although the two problems were formally similar.

**12.** Results concerning the implicit learning of invariant features (see note 10) are more complex because positive transfer was initially reported (e.g., McGeorge & Burton 1990). However, Stadler et al. (2000) have shown that transfer disappears when the response strategy initially discovered by Wright and Burton (1995) is denied to subjects. They concluded that "this form of learning, like many other forms of implicit learning and memory, is hyperspecific" (Stadler et al. 2000, p. 235).

## APPENDIX A: PARSER

PARSER is centered on a single vector, called percept shaper (PS). At the start, PS contains only the primitives composing the material, namely, a few syllables. Learning proceeds through the iterative processing of small parts of the linguistic corpus which can be in immediate succession or separated by a various amount of unprocessed material according to the simulations. Each part is composed of one to three processing primitives (the number is determined randomly for each percept), thus simulating the successive attentional focuses of a human subject processing the same corpus. Each perceived part is added to PS, and can itself serve as a new primitive for the shaping of subsequent inputs, as the syllables initially did. This simulates the fact that perceptual contents are changing throughout the task. Finally, if learning has been successful, PS contains all the words, and only the words of the language.

Why does PS not become encumbered with an innumerable set of irrelevant and increasingly lengthy units? It is because the future of a unit depends on its weight, which represents trace strength. The weight of a given unit is incremented each time this unit is perceived (weight = + 1), and decremented each time another unit is perceived (decrement = −0.05). Decrement simulates forgetting. In order to fulfill its shaping function, any unit of PS needs to reach a threshold value (threshold = 1). As a consequence, a unit needs to be perceived repeatedly and regularly in order to persist on fulfilling a shaping function. In contrast, when the frequency of perception of a given element is not high enough to counteract the effects of forgetting, this element is removed from PS when its weight becomes zero.

It must be understood that the details of the functioning of the model are not intended to provide a realistic picture of the processes that are actually involved. As a case in point, forgetting is simulated through the linear decrement of a weight, whereas there is evidence that the forgetting curve fits only moderately well with a linear trend. Needless to say, whatever the mathematical function, forgetting is certainly not biologically implemented as the decrement of a numerical value. We believe the use of this artificial mean unimportant, given that the simulated result, forgetting, corresponds to an ubiquitous phenomenon.

More important, it may be argued that the general architecture of PARSER is not compatible with the mentalistic view. Indeed, PS may be thought of as a memory store or a mental lexicon, in which symbolic representations exist independently of the current phenomenal experience of the subject. This possibility is not actually permitted in our general framework. Still more worrying is the fact that the items in PS with a weight lower than one could be viewed as instantiating "deeply unconscious representations," since they are stored without being able to shape the content of the phenomenal experience.

The contradiction is indeed evident, but, we believe, not detrimental to the demonstration provided by PARSER of the power of the general principles it implements. Indeed, the representations stored in PS, whatever their weight, play a role only when they match the external input. They perform no function except when they enter as a component of the current phenomenal experience. As argued in the main text (see Note 1), the same result should have been obtained had the memory of the system been simulated as a capacity to build an on-line representation in the presence of a given input, without directly storing the representation itself. Low-weighted items could have been replaced by a procedure in which, instead of creating new traces ex nihilo, a given input excites the same processing path as a prior, identical input, thus reinforcing this processing path. High-weighted items could have been replaced by a procedure in which the path excited by the processing of these items is strong enough to guide the formation of the current percept.

In fact, neural network modeling would certainly have been more in keeping with our approach, because it naturally implements the idea that the memory of the system is not necessarily a list of symbolic tokens. However, in most connectionist models,



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

the representations embedded in the connection weights between units are not formatted to serve as new coding primitives. This makes it difficult to implement the idea that associations apply to increasingly complex representations. However, this is a technical difficulty that does not require us to discard connectionist modeling altogether. For instance, the Cascade Correlation architecture (Fahlman & Lebiere 1990) may help solve the problem, due to its ability to dynamically build high-order feature detectors by adding extra units to small networks. Such algorithms, or other so-called constructive methods, seem to represent a promising way to implement the principles underlying PARSER in the form of a connectionist network.

ACKNOWLEDGMENTS

This research was supported by the Centre National de la Recherche Scientifique (UMR CNRS 5022), the University of Bourgogne, and the Region of Bourgogne (Contract AAFE).

# Open Peer Commentary

*Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.*

## Is the self-organizing consciousness framework compatible with human deductive reasoning?

Pierre Barrouillet[a] and Henry Markovits[b]

[a]*Université de Bourgogne, LEAD-CNRS, Pôle AAFE, Dijon, France;* [b]*Département de psychologie, Université du Québec à Montréal, Montréal, QC, Canada.* **barouil@u-bourgogne.fr** **markovits.henry@uqam.ca**

**Abstract:** As stressed by Perruchet & Vinter, the SOC model echoes Johnson-Laird's mental model theory. Indeed, the latter rejects rule-based processing and assumes that reasoning is achieved through the manipulation of conscious representations. However, the mental model theory as well as its modified versions resorts to the abstraction of complex schemas and some form of implicit logic that seems incompatible with the SOC approach.

Deductive reasoning is one of the main human activities for which a sophisticated and powerful unconscious has often been advocated. For example, so-called "mental logic" theories assume that human beings possess unconscious rules of reasoning that are isomorphic to some of the rules of formal logic (Braine & O'Brien 1998; Rips 1994). These rules are applied automatically when their conditions of application are met, resulting in the production of a conclusion, which is the only conscious trace of this process. It has also been claimed that the selection of information from which conclusions are reached, escapes consciousness. For example, Evans (1989) distinguished between heuristic processes aimed at selecting psychologically relevant information and analytic processes intended to reach some conclusion from this information. He stressed the pre-attentional and unconscious character of the heuristic phase, which would be directed by cognitive biases and implicit knowledge that determine relevance. The recourse to unconscious heuristic processes as an explanatory device not only accounts for reasoning errors, but also implies that any attempt to understand reasoning through explicit justifications and verbal reports is fruitless; because the *reasons* for a given conclusion necessarily remain out of the reach of consciousness. According to this view, justifications are just ad hoc rationalizations unrelated to the reasoning process. More recently, Oaksford and Chater (2001) suggested that reasoning is based on probabilities computation. For example, they assume that when solving Wason's selection task, individuals calculate probabilities using Bayes' theorem. It can be inferred that these computations are unconscious, particularly because few individuals can successfully perform these complex calculations when explicitly asked to do so.

Thus, the conventional cognitive framework mainly considers reasoning to be a series of unconscious processes that produce conclusions that would appear, from a phenomenological point of view, to be sprung from nowhere. This is in sharp contrast to the commonsense viewpoint, which considers deductive reasoning to be a highly conscious, accessible, and declarative process, the steps of which cannot only be reported, but also justified. However, does the SOC model constitute an alternative to the conventional cognitive framework in understanding human deductive reasoning? Prima facie, the answer seems to be positive, but a closer scrutiny could lead to some doubts.

The mental model theory of deductive reasoning has already argued that people reason from representations that are isomorphic to the structure of the world, and that this isomorphism may underpin forms of logical reasoning seemingly rule-governed (Johnson-Laird 1983; Johnson-Laird & Byrne 1991). Thus, Perruchet & Vinter naturally evoke mental models as an example of theory that rejects abstract computation and rule-based processing. Indeed, according to Johnson-Laird, manipulating transient (and conscious) mental models whose structure is isomorphic of the state of affairs the given premises describe, could lead to logically correct conclusions by processes that obviate the need for recourse to logical rules. However, are the mental models for reasoning entirely comparable with the SOC representations resulting from associative learning and memory? Not exactly. Indeed, deductive reasoning by mental models actually involves complex formal structures for representations as well as logical principles for coordinating and manipulating mental models that are very close to logical rules. For example, Johnson-Laird and Byrne (2002) assume that there is a "core meaning" for the conditional (i.e., sentences of the form "if $p$ then $q$") of the form

$$p\ q$$
$$\neg p\ q$$
$$\neg p\ \neg q$$

in which each line represents a different state of affairs with the sign $\neg$ for the negation. It is noteworthy that this core meaning is a kind of abstract and formal schema that needs to be instantiated by specific values. Though the SOC theory assumes that representations can be schematic, flexible, and prone to generalization, it is not clear that associative learning processes are sufficient to extract such a complex abstract representation from attended regularities of experienced situations to which conditional verbal inputs apply.

A further problem is raised by the processes used to coordinate and simplify models, which are very akin to logical rules. For example, contradictory models that contain both $p$ and $\neg p$ would be eliminated following a heuristic that closely resembles the rules that define contradiction in mental logic theories (e.g., Braine & O'Brien 1998). The coordination of different models is achieved through a concatenation process formally equivalent to the logical rule "*and* introduction" in mental logic models (Braine & O'Brien 1998; Rips 1994). In other words, the standard mental model theory for deductive reasoning resorts to representational structures and cognitive processes that exceed the narrower scope of associative learning and memory.

Of course, attempts have been made to avoid the formal aspects of this standard theory. For example, we have recently assumed that the mental models for conditional reasoning do not correspond to any formal schema or core meaning (Markovits & Barrouillet 2002a; 2002b). Instead, they would reflect the structure of knowledge that individuals have gathered about the relation in-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

volved in the conditional sentence. Thus, mental models would not represent a list of possible co-occurrences but rather, meaningful relationships between specific values or events that are retrieved from long-term memory. These assumptions are closer to the SOC theoretical framework while accounting for phenomena usually ascribed to a complex and unconscious computational mechanism. For example, our model mimics content effects by producing conclusions that are sensitive to the relative probabilities of occurrence of the relevant events in the real world. However, the probabilities involved in our computational model are not *computed* by the cognitive system but are simply inherent properties of the processes of memory retrieval used to produce the conscious representations that are used in producing a conclusion.

While this approach can provide a model of reasoning with concrete premises that is more compatible with the SOC framework than is standard mental model theory, one critical exception concerns the late-developing, and relatively rare, capacity to reason logically with unfamiliar and even abstract material. Venet and Markovits (2001) found that abstract reasoning is very difficult for younger adolescents who can reason very well with concrete premises, and that abstract reasoning does not appear to be explicable by the same kinds of cognitive processes used in concrete reasoning. They suggested that the basic processes used in abstract reasoning differ qualitatively from those used in concrete reasoning and are the result of the kind of representational redescription described by Karmiloff-Smith (1995). This would imply that individuals use their experience with concrete forms of reasoning to develop relatively abstract representations of the basic types of information that characterize conditional reasoning. Such a process would in fact lead to exactly the kind of definition of the conditional that has been postulated by Johnson-Laird, that is, a formal schema that includes a complex logical apparatus with propositional-like tags for negation and mental footnotes for exhaustion, as well as logical rule-like processes for manipulating representations.

Thus, even if at a first sight it seems easy to account for human deductive reasoning without a sophisticated and logical cognitive unconscious, things are more complicated. Though rejecting abstract computation and rule-based processing, both Johnson-Laird's model and our modified mental model theory are simply unable to account for all aspects of human deductive reasoning without any recourse to abstraction processes of complex schemas and some form of implicit logic. As stressed by Perruchet & Vinter themselves, the SOC interpretation of abstraction "is viable only if the coding of the incoming information in an abstract and relational format remains simple enough to be attributed to low-level perceptual processes" (target article, sect. 6.2). We can doubt that this is the case for a complex logical structure such as the conditional.

## Varieties of consciousness

Paolo Bartolomeo and Gianfranco Dalla Barba

*INSERM, Centre Paul Broca, Paris, F-75014, France.*
**paolo@broca.inserm.fr dallabarba@broca.inserm.fr**
**http://paolo.broca.inserm.fr**

**Abstract:** In agreement with some of the ideas expressed by Perruchet & Vinter (P&V), we believe that some phenomena hitherto attributed to "unconscious" processing may in fact reflect a fundamental distinction between direct and reflexive forms of consciousness. This dichotomy, developed by the phenomenological tradition, is substantiated by examples coming from experimental psychology and lesion neuropsychology.

Perruchet & Vinter (P&V) have made a convincing case for a role of consciousness in phenomena hitherto supposed to be examples of unconscious processing, such as implicit learning or semantic priming. However, in some of these cases people are not willing to acknowledge that they had experienced the relevant stimuli (e.g., rules or primes), thus suggesting "unconscious" processing of these items. But does a lack of verbal report necessarily indicate unconscious processing? Whereas an appropriate verbalization can be considered as a reliable indicator of conscious processing (Merikle et al. 2001), the converse is not necessarily true; lack of verbalization cannot be conclusively considered to indicate lack of consciousness. For example, it might simply indicate lack of memory (Allport 1988). The phenomenological tradition has often distinguished between direct and reflexive forms of consciousness (review in Vermersch 2000; see Marcel 1988 and Dulany 1997 for more recent proposals of similar dichotomies). This distinction may help explain why people observing an array of letters for a very short time may be aware of having seen letters but able to name only a subset of them (Sperling 1960). In the same vein, studying the so-called endogenous mode of orienting of spatial attention, usually attributed to voluntary processes (Jonides 1981; Posner & Snyder 1975), Decaix et al. (in press) recently found that participants were able to develop effective strategies despite the absence of explicit instructions. About half of participants, however, were later unable to correctly describe the strategy they used, even when this description was proposed in a post-experiment questionnaire. These examples may represent forms of direct consciousness that, for various reasons, cannot translate in more reflexive forms of consciousness. To borrow the terms used by Merleau-Ponty (1942), one can "live" forms of perception that one cannot speak about. Merleau-Ponty uses the example of someone who enters a room and feels an impression of disorder, only to later discover that this impression came from a crooked picture on the wall. Before discovering that, this person's consciousness was "living things that it could not spell out." This would by no means imply that the first impression on entering the room was *unconscious*! Rather, it was a form of consciousness not immediately amenable to verbal description.

Neuropsychological evidence from brain-damaged patients offers instances of the opposite dissociation, with defective direct consciousness and preserved reflexive consciousness. Patients with left unilateral neglect typically lack phenomenal awareness for events occurring in the neglected part of space, perhaps because these events fail to capture their attention (see Bartolomeo & Chokron 2002 for a recent review). Although in general these patients are reluctant to acknowledge their disorder, some eventually become cognizant of their neglect, but continue nevertheless to show the symptom. A celebrated film director F.F. jokingly asked to include his new condition of neglect in his calling card, but persisted in producing funny drawings lacking their left part (Cantagallo & Della Sala 1998). In another cognitive domain, we described a patient with a severe deficit of so-called frontal cognitive abilities (such as planning and memory-related activities) (Dalla Barba et al. 1999). Despite his unawareness of the deficits, this patient was surprisingly aware of his incapacity to appreciate his disorder. The following excerpt from an interview with the patient describes well the dissociation between impaired direct consciousness and preserved reflexive consciousness.

Q. How is your memory?
A. Good.
Q. How do you think other people judge your memory?
A. Bad.
Q. Do you think they are right or wrong?
A. Right.
Q. Why?
A. Because they see what I don't see. Yes, it's precisely that. I don't see what they see. They see a problem I don't see.
Q. So, at the same time, you are conscious of your memory being good and bad. Isn't this quite a bizarre situation?
A. Well, it is, but it's just like that. When I think about my memory, or simply of my ability to do things, I think I am completely normal. But then, when I think how my wife or my eight-year-old kid looks at me and reacts to my behavior, I realize that I am not aware of something in my behavior that is wrong, that I can't see and that they see.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Q. And you think they are right?
A. Yes, they are right.
Q. Why are you so sure? They could be wrong.
A. Well, it's me who had a head trauma, not them.

Also in this case, unfortunately, this reflexive awareness was ineffective to palliate the patient's cognitive problems.

These examples suggest the opportunity of considering that different varieties of consciousness may be involved in human behavior, before invoking unconscious processes as a default explanation for dissociations between what in fact may be direct and reflexive forms of consciousness. We believe that these considerations may inspire promising new avenues of research in cognitive neuroscience, which integrate the methods of experimental psychology with the insights about the taxonomy and the operations of consciousness coming from the phenomenological tradition (Dalla Barba 2002).

## Consciousness organizes more than itself: Findings from subliminal mere exposure research

Robert F. Bornstein
*Department of Psychology, Gettysburg College, Gettysburg, PA 17325.*
**bbornste@gettysburg.edu**

**Abstract:** Contrary to Perruchet & Vinter's self-organizing consciousness (SOC) model, subliminal mere exposure (SME) research indicates that stimuli perceived without awareness produce robust effects. Moreover, SME effects are significantly stronger than mere exposure effects produced by clearly recognized stimuli. The SOC model must be revised to accommodate findings from studies that use affect-based outcome measures.

After reviewing empirical research on the cognitive unconscious, Perruchet & Vinter (P&V) present a simple but powerful thesis: that the end products of conscious phenomenal experience are rooted in successive iterations and articulations of simpler conscious experiences. P&V explore these processes within the context of an innovative conceptual framework they call *self-organizing consciousness* (SOC). They contend that SOC can account for a broad array of ostensibly "unconscious" phenomena (e.g., problem solving, implicit learning) without reference to mental processing outside of awareness.

Although P&V noted the potential importance of emotionally valenced stimuli within the SOC framework, their analysis is limited primarily to studies that use affectively neutral stimuli and outcome measures. Moreover, their review of preference-based decision making (sect. 7.2) – a potential bridge between affectless and affect-driven cognitive processes – excludes a number of seminal studies. Although P&V reviewed some key investigations of implicit learning and implicit memory (sect. 8.1), here too they focused on those studies that employed affectively neutral stimuli (i.e., repetition priming studies) and ignored studies wherein the effects of repeated stimulus presentations are assessed via affect-based outcome measures.

P&V acknowledge that they might have inadvertently excluded certain phenomena that could compel them to revise the SOC model (sect. 9.2). The purpose of this commentary is to describe one such phenomenon: The subliminal mere exposure (SME) effect, a well-established implicit memory task involving preference-based decisions.

The SME effect fits P&V's rigorous criteria for unconscious mental processing (sect. 8), as well as other widely used criteria for demonstrating perception and memory without awareness (Bornstein 1992). A detailed discussion of SME research is beyond the scope of this commentary (see Zajonc 2001 for a recent review). In the context of P&V's SOC framework, six issues are germane:

1. *SME effects are robust and replicable.* To date, more than 20 published experiments have obtained robust SME effects (Bornstein 1992; Zajonc 2001). These experiments have come from at least seven independent laboratories, and involved a variety of stimuli (e.g., photographs, polygons, ideographs, nonsense words, musical selections, line drawings, impossible objects), and outcome measures (e.g., forced-choice preference judgments, liking ratings, pleasantness ratings, behavioral preference indices).

2. *SME effects are obtained with very brief stimuli.* Researchers have obtained SME effects using exposure durations as brief as 1msec, with stimulus afterimage confounds controlled through use of a post-stimulus pattern or energy mask (Bonnano & Stillings 1986; Kunst-Wilson & Zajonc 1980).

3. *SME effects are obtained even when objective awareness indices are used.* Although a number of procedural variations have been used, the typical SME experiment requires participants to make forced-choice preference and recognition judgments for previously seen and novel stimuli, with order of judgments counterbalanced across participants. In every published SME experiment to date, recognition judgments have not exceeded chance performance (i.e., 50% accuracy), even when previously seen stimuli are preferred over novel stimuli at better-than-chance levels. Some experiments (e.g., Bornstein & D'Agostino 1992) have further demonstrated that participants in an ancillary stimulus discrimination task cannot distinguish stimuli from blank slides under conditions identical to those in the experiment proper.

4. *SME effects persist over time.* Only one investigation has explored the temporal duration of SME effects, but the results of this study confirmed that these effects persist undiminished for one week following stimulus exposures (Seamon et al. 1983).

5. *SME effects generalize to similar stimuli.* As P&V note (sect. 8.2.1), Bar and Biederman (1998) found no transfer effects in a forced-choice recognition task following subliminal stimulus exposures. This is the typical result obtained in SME research as well. However, when preference judgments are used in lieu of recognition judgments, robust transfer effects are observed, with stimuli that are conceptually or structurally similar to previously exposed stimuli preferred over stimuli that are less similar (Zajonc 2001).

6. *SME effects are significantly stronger than mere exposure effects produced by stimuli that are consciously perceived.* In a comprehensive meta-analysis of the SME literature, Bornstein (1989) found that the magnitude of the SME effect is several times greater than the magnitude of the mere exposure effect produced by stimuli that are consciously recognized. The overall magnitude of the exposure effect (represented by the effect size indicator *r*) was .528 for stimuli that were not recognized at better-than-chance levels, and .121 for stimuli that were recognized at close to 100% accuracy. Bornstein and D'Agostino (1992) confirmed these meta-analytic results in two experiments comparing the magnitude of the exposure effect produced by identical unrecognized and clearly recognized stimuli.

SOC is a useful model for certain cognitive processes, but it cannot account for the entire range of human phenomenal experience, nor render obsolete the concept of the cognitive unconscious. SME effects are particularly problematic for the SOC model, because not only do stimuli perceived without awareness produce robust changes in affective responding, but these changes are significantly *greater* than those produced by identical stimuli that are consciously perceived. These results do not suggest that the unconscious is "omnipotent" (sect. 1), but they do imply that unconscious and conscious mental processes involve different affective dynamics.

In light of findings involving the SME effect, P&V's SOC framework must either be: (a) revised, to accommodate the role of implicit processes in preference-based decisions; or (b) limited in scope, so it applies only to those psychological phenomena wherein effects are assessed via cognitive (not affect-based) outcome measures. Future studies will determine which of these two paths SOC must take, but two decades of SME research confirm


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

that stimuli perceived and remembered without awareness have robust effects on human behavior and affective responding.

## Mentalism, information, and consciousness

Richard A. Carlson

*Department of Psychology, The Pennsylvania State University, University Park, PA 16802-3106.* **racarlson@psu.edu**
**http://psych.la.psu.edu/faculty/carlson.htm**

**Abstract:** The target article addresses important empirical issues, but adopts a nonanalytic stance toward consciousness and presents the mentalistic view as a very radical position that rules out informational description of anything other than conscious mental states. A better mentalistic strategy is to show how the structure of some informational states is both constitutive of consciousness and necessary for psychological functions.

Perruchet & Vinter (P&V) make two important claims: that learning results only from operations on representational elements jointly active in awareness, and that empirical evidence for unconscious mental contents is flawed. These conclusions are said to support the central mentalistic tenet "that the only cognitive representations are those that form the phenomenal experience" (sect. 8.2.1). I am sympathetic to each of these points. The problem is that cognitive science has had substantial success in using informational descriptions of many processes that are clearly not conscious, and "representation" serves as a broad category term for states on which these processes operate. This success licenses the metatheory on which only some representations are conscious, leading to the epiphenomenalism about consciousness that P&V (and I) reject. It is important to demonstrate that contents of conscious states can account for phenomena presumed to require unconscious mental contents, but such demonstrations are unlikely to settle the broader issue. A proponent of the mentalistic view thus has two choices: to argue that describing nonconscious states in informational terms is inappropriate or (merely) metaphorical, or to suggest principles for distinguishing the informational states that constitute conscious experience from those that do not. P&V apparently prefer the first choice, suggesting that "the notion of computation does not extend to any neural activity." (sect. 1.3.2). This position will seem untenable to many cognitive scientists, and I think that the issue between mentalism and the cognitive unconscious is better cast in terms of the second choice.

As I see it, the issue is whether we need theories that postulate states that are essentially *like* conscious mental states but not part of the stream of subjective experience. For example, we know what it is like to hold a rule (consciously) in mind and use it to control our behavior. Are there closely analogous processes that involve similar representations but are unconscious? The mentalistic position is that no states like conscious mental states (but unconscious) need be invoked to explain psychological phenomena.

Respecting P&V's restriction of the term *representation* to contents of awareness, let's call the larger category *informational states* – states whose roles in cognition and behavior are by virtue of having an appropriate informational description. To locate consciousness within the general computational framework, we need to understand how those informational states that constitute states of awareness differ from those that don't. It can't be just that they are conscious – that way lies the epiphenomenalism about consciousness that P&V (and I) reject. Could it be that they are the ones that serve "the function of some meaningful component of the represented world . . . within the representing world" (sect. 1.3.2)? That is more promising, but a lot rides on what constitutes a meaningful component, and to avoid circularity it can't be that the meaningful components of the world are just those of which an individual is aware. Could it be that the informational states that constitute conscious states are the ones that "may also enter as an argument into deliberative mental episodes" (sect. 1.3.2)? Again, to avoid both circularity and epiphenomenalism, it can't just be that such episodes are conscious, thus conferring consciousness on the participating informational states. How can we break out of this circle?

Here's what I think: Some informational states have a characteristic structure that both allows them to serve certain mental functions, and that is constitutive of consciousness. But when we analyze the informational state that constitutes a particular, momentary, conscious mental state, we'll find that only some aspects of that informational state constitute the *content* of awareness in that state. Elsewhere, theorists with mentalistic inclinations such as myself (Carlson 1997; 2002), Dulany (e.g., 1968), and Zelazo (e.g., 1999) have offered specific proposals about how to analyze conscious mental states and their roles, in particular, psychological processes such as learning and control by rules. In each of these theories, content is only one aspect of a conscious state. And it is by virtue of the structure that makes an informational state an instance of a conscious mental state that it can serve particular cognitive functions. For example, Zelazo (1999) argues that recursive awareness of an object in terms of a semantic description is necessary to support action toward that object when it is not perceptually present.

P&V might or might not agree with any of these analyses, but to link their conclusions about empirical phenomena to the general debate about a cognitive unconscious requires some story about which informational states constitute conscious states, and how they do so. Alternatively, I think they are stuck with a very radical view that rules out informational description of anything other than conscious mental states. P&V cite O'Brien and Opie (1999a) for "a link between consciousness and representation," but a central feature of that article is the authors' effort to specify what it is about a pattern of neural activity (an informational state) that makes it a representation and allows it to serve particular information-processing functions.

A further difficulty stemming from P&V's nonanalytic stance toward consciousness is unfortunately characteristic of much research they cite – it is difficult to know exactly what hypotheses are being proposed or rejected. For example, theorists who argue that what makes a representation conscious is having a higher order thought about it (e.g., Rosenthal 1993) might argue that the kind of minimal representation in awareness proposed by P&V is, in fact, unconscious representation. To others it will seem like a clear example of what Zelazo (1999) describes as "minimal consciousness." And many of those who argued that classical or operant conditioning could occur without awareness meant that individuals might learn without becoming aware of stimulus-stimulus or stimulus-response *contingencies* that defined the learning situation (Brewer 1974), a possibility that seems to be allowed by P&V. Of course, as P&V note, some have taken evidence that particular phenomena involve only minimal consciousness as evidence for an extensive, powerful, unconscious cognitive system that is like conscious thinking but operates "outside" awareness (e.g., Erdelyi 1992).

There is much to be said for a mentalistic view that establishes a special explanatory role for consciousness. Challenging evidence said to require a cognitive unconscious is part of the job, but we also need theories of consciousness sufficiently powerful to distinguish conscious mental states from other kinds of informational states. What a modern mentalist ought to want, I think, is exactly a theoretical reason that some informational states both are conscious and serve particular psychological functions by virtue of being conscious.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

## Modeling consciousness

Frédéric Dandurand and Thomas R. Shultz

*Department of Psychology, McGill University, Montréal, Québec, H3A 1B1 Canada.* **F.Dandurand@sympatico.ca thomas.shultz@mcgill.ca**
**http://www.psych.mcgill.ca/perpg/fac/shultz/default.html**

**Abstract:** Perruchet & Vinter do not fully resolve issues about the role of consciousness and the unconscious in cognition and learning, and it is doubtful that consciousness has been computationally implemented. The cascade-correlation (CC) connectionist model develops high-order feature detectors as it learns a problem. We describe an extension, knowledge-based cascade-correlation (KBCC), that uses knowledge to learn in a hierarchical fashion.

Issues concerning the role of consciousness in cognition and learning have bedeviled psychology for many years. It has proved to be enormously difficult to delineate the role of consciousness in various psychological phenomena. Perruchet & Vinter (P&V) show that a theory emphasizing the unconscious processing of information about which people are conscious accounts for a wide range of phenomena just as well as a theory that allows for unconscious processing of information about which people are not conscious. Or, put more simply, assuming conscious awareness of information fits the data just as well as assuming an unconscious does. Although this is useful and even interesting, it does little to resolve the classical issues of the roles of consciousness and the unconscious in cognition and learning. The idea that multiple theories account for a range of phenomena is common in psychology and remains true in this domain.

With their PARSER simulations, P&V bring a potentially useful tool to the arena. The critical importance of attentional focus in this model is extremely interesting. As P&V correctly note, however, computational models are essentially neutral with respect to consciousness because none of them actually have it. An important reason why none of them have it is that no one yet knows enough about consciousness to implement it computationally. Equating consciousness with attention, as in PARSER, is an interesting gambit, but it is far from clear that consciousness has been fully and effectively implemented in PARSER. Many researchers would insist on some level of awareness as a key marker for consciousness.

We note with interest that P&V intend to implement PARSER in a connectionist framework, specifically with CC algorithm (Fahlman & Lebiere 1990), which we have used in a wide variety of psychological simulations (Shultz, in press). A recent simulation concerned a data set given extensive coverage by P&V, that of Marcus et al. (1999). This simulation showed that a cascade-correlation encoder network covers the major features of Marcus et al.'s infant data (Shultz & Bale 2001) and generalizes outside the range of the training patterns, at the same time showing psychologically plausible content effects. Like most connectionist models, this one requires neither conscious awareness nor explicitly built-in attentional focus. Attention is not a cause of the model's successful learning but a result of learning – the networks learn what features of the stimuli to attend to.

Perruchet and Vinter foresee two problems in implementing PARSER using CC: (1) the representations embedded in the connection weights between units are not formatted to serve as new coding primitives; and (2) it is difficult to implement the idea that associations apply to increasingly complex representations. Despite the fact that CC develops high-order feature detectors as it learns a problem, P&V correctly point out that explicit associations between simple and complex representations are not modeled using standard CC. Our extension of CC (KBCC) explicitly addresses the problem of knowledge re-use and the building of complex representations (Rivest & Shultz 2002; Shultz & Rivest 2001; Thivierge & Shultz 2002).

Both KBCC and CC are constructive methods in which the network topology expands as necessary for learning the problem at hand. Expansion occurs when a new unit from a pool of candidates is added into the network. The main improvement of KBCC over CC is in the content of the pool of candidates used for recruitment. Whereas CC uses only simple units, KBCC extends the pool of candidates to arbitrarily complex CC networks. Knowledge contained in previously trained CC networks is therefore directly available for use in new problems, hence implementing knowledge re-use and solving the first issue raised by P&V.

Although KBCC can recruit any kind of sub-network if it correlates best with the residual network error, there are three particularly useful sources of knowledge: sub-tasks (e.g., rules for DNA splicing; Thivierge & Shultz 2002), simple components of a task (e.g., vertical and horizontal components of a cross; Shultz & Rivest 2001), and analogous tasks (e.g., vowel recognition of male speakers as relevant knowledge for recognizing female speakers; Rivest & Shultz 2002).

Because CC and KBCC networks start with little internal structure, they tend to learn simple ideas first and build more complex ideas on top of simpler ones, hence learning hierarchically. Associations between simple and complex representations are implemented in the connection weights between the network learning the complex task and a recruited sub-network having learned a simpler task. This addresses the second issue of P&V. Additionally, KBCC creates new connections when recruiting a sub-network into the current network topology. Only those new connections are trained, and therefore knowledge of the simpler task embedded in the recruited network is left intact.

There are two differences between KBCC and PARSER that should be further investigated. First, KBCC is a supervised learning algorithm, as opposed to PARSER, which is unsupervised. Using a variant of CC called recurrent cascade-correlation (RCC) (Fahlman 1991), it is possible to define the supervised task as the prediction of the next symbol in the input sequence, hence solving the problem of providing a target vector. Alternatively, if the task can be construed as an encoder problem, in which stimulus patterns are encoded onto hidden representations and then decoded onto output units, the need for a separate target vector also evaporates.

Second, KBCC does not learn all levels of a complex task as seamlessly as Perruchet and Vinter are aiming for. Instead, KBCC requires external intervention to present tasks in a timely and ordered fashion. After each sub-task is learned, KBCC is restarted with the next task as a goal and the networks trained on previous tasks as part of its candidate pool.

Nonetheless, KBCC does address the issues raised by P&V. Within KBCC, source networks are naturally formatted to serve as new coding primitives. This allows knowledge encoded in the connection weights of trained networks to be re-used to solve more complex problems. Furthermore, prior knowledge of simple tasks remains intact after recruitment, and new connection weights to the recruited network implement the association between simple and complex representations.

ACKNOWLEDGMENTS

Our research is supported by grants from the Natural Sciences and Engineering Research Council of Canada and from Quebec FCAR.

## The self-organizing conundrum

Arnaud Destrebecqz and Axel Cleeremans

*Cognitive Science Research Unit, Université Libre de Bruxelles, Fonds National de la Recherche Scientifique, B-1050, Belgium.* **adestre@ulb.ac.be axcleer@ulb.ac.be http://srsc.ulb.ac.be/axcWWW/axc.html**

**Abstract:** Perruchet and Vinter stop short of fully embracing the implications of their own SOC framework, and hence end up defending an implausible perspective on consciousness. We suggest instead that consciousness should be viewed as a graded dimension defined over quality of representation. This graded perspective eliminates the most problematic aspects of the cognitive unconscious without denying its existence altogether.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Perruchet & Vinter (P&V), like others before them (e.g., Cleeremans 1997; Cleeremans & Jimenez 2002; O'Brien & Opie 1999a; 1999b; Searle 1992), rightfully reject the "classical" notion that cognition exclusively involves symbol manipulation. The specific problem that computational perspectives of cognition have with consciousness is this: Representations in such systems, because they consist of nothing else than structured propositions stored in mental databases, are causally inert and hence cannot influence processing in any way that does not involve "access" by an external agent. However, the possibility of such influence is exactly what one observes in cases where the contents of awareness appear to be dissociated from action, which, remarkably, even critics like P&V recognize exist.

There are essentially three strategies to deal with implicit cognition – two bad ones, and one that might just end up working. The first strategy consists of ascribing the phenomena of implicit cognition to an altogether separate system. This is what one could call the "Zombie" strategy. The second strategy consists of denying implicit cognition altogether. This is what we call the "Commander Data" strategy: Cognition is fully transparent to yourself. The third strategy consists of abandoning the idea that cognition exclusively involves symbol manipulation. This is the perspective that P&V embrace. Yet their proposal ultimately turns out to be unsatisfactory, essentially because they stop short of exploring the full implications of a genuinely nonsymbolic perspective.

In this context, it is deeply ironic, as P&V admit themselves, that PARSER ends up conjuring exactly the sort of characterization of the relationships between the conscious and the unconscious that the authors otherwise so strongly reject: that of the continued existence of subthreshold yet causally efficacious representations. It is crucial to realize that subthreshold units in PARSER, far from being impotent, are in fact central to its functioning. To see this, one needs only to consider that a unit that is just below threshold has a greater probability to shape processing on the next episode than low-strength units. Subthreshold cognition is thus functional after all, if only potentially so. P&V dodge this problem by urging us to think of subthreshold representations not as representations at all, but rather as something similar to the connection weights of a neural network. P&V further suggest that their ideas would in fact be better embodied by constructivist connectionist networks.

Let us then consider what it would mean to think of PARSER as a connectionist network. Two central features of such models, which we view as essential to our understanding of the relationships between conscious and unconscious processing, are: (1) that processing has a inherently graded character; and (2) that representations are constantly causally efficacious because they arise as a mandatory consequence of processing, by virtue of being embedded in a causal chain. To capture these properties through PARSER, one might therefore wish to relax the constraint of enforcing a strict, arbitrary threshold to separate functional and nonfunctional representations. Thus, instead of assuming that only suprathreshold representations shape perception, assume instead that each unit can shape perception with a probability that depends on its strength. This would presumably make perception somewhat less efficient, which, as it turns out, would not be a bad thing at all, because PARSER is in fact far too powerful in its current incarnation (as noted by the authors). This simple change, however, weakens the SOC proposal considerably, for we now have low-strength representational objects that can potentially shape perception. In other words, we now have *graded* representations, the weak instances of which one could choose to describe as part of a "fringe" of consciousness as unconscious representations, or transient representations.

It is important to realize that such weak representations occur constantly in interactive neural networks in the form of patterns of activation over given processing units. They are produced by information contained in connection weights, which, while they cannot be accessed as representations by the network itself, nevertheless drive processing by generating patterns of activation over the units that they are connected to. The crucial point is that in such networks, patterns of activation arise as a mandatory consequence of processing. They might be weak or strong, specific to a given stimulus or not, stable in time or less so. We take these different aspects to correlate with a representation's availability to both phenomenal experience and to control processes.

Embracing the idea that representations may be graded opens up many interesting conceptual possibilities. First, gradedness over representations makes it possible to understand why phenomenal experience also seems to have a graded character. Attention selects the strongest representations that are active at some point in time to form the focus of awareness, but there is also clear first-person evidence that our phenomenal experience is in fact far richer and includes a fringe of attenuated contents.

Second, gradedness makes it possible to situate both implicit cognition and automaticity on a single continuum defined by quality of representation. From this perspective (developed in Cleeremans & Jiménez 2002), implicit cognition involves weak, poor-quality mental representations that are (because of their weakness) both difficult to control (Destrebecqz & Cleeremans 2001) and only weakly available to phenomenal experience. Crucially, such representations can nevertheless influence behavior, for example, through processes of priming that involve multiple weak representations. Automaticity, in contrast, involves very strong representations, which, while available to conscious awareness, are difficult to control (because of their strength).

The Stroop situation is a perfect example of such cases, where one experiences the conflict yet remains incapable of exerting sufficient control to resolve it. It is worth pointing out here that as long as one assumes that strong representations emerge through processes of adaptation, one can indeed conclude, just as P&V did, that learning shapes conscious experience, and that conscious experience, in turn, shapes learning: Strong representations are adapted because the learning processes that produce them are adaptive themselves. Consciousness is thus self-organizing indeed, but one has to start somewhere, however, and this starting point consists of the weak representations characteristic of implicit cognition.

This is the conundrum, then: Developing the implications of the SOC framework to their fullest extent inevitably leads one to reject the very premises on which it is based, and to accept instead that consciousness is a graded dimension. It then follows that one can both reject the notion of a strong, symbolic unconscious, while at the same time recognizing that some (weak) representations are best described as unconscious. From this perspective, the challenge is therefore not so much to explain cognition by appealing exclusively to conscious representations, but rather to explain consciousness by figuring out which aspects of information processing determine why some representations become conscious and others do not.

ACKNOWLEDGMENTS

Arnaud Destrebecqz is a Postdoctoral Researcher with the Fonds National de la Recherche Scientifique (F.N.R.S., Belgium). Axel Cleeremans is a Senior Research Associate with the F.N.R.S. This work was supported by a grant from the Université Libre de Bruxelles in support of IUAP program P4/19 and by grant HPRN-CT-2000-00065 from the European Commission.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

## What sort of representation is conscious?

Zoltan Dienes[a] and Josef Perner[b]

[a]*Department of Experimental Psychology, Sussex University, Brighton, BN1 9QG, United Kingdom;* [b]*Department of Psychology, University of Salzburg, A-5020 Salzburg, Austria.* **dienes@biols.susx.ac.uk**
**josef.perner@sbg.ac.at**
**http://www.biols.susx.ac.uk/home/Zoltan_Dienes/**
**http://www.sbg.ac.at/psy/people/perner_e.htm**

**Abstract:** We consider Perruchet & Vinter's (P&V's) central claim that all mental representations are conscious. P&V require some way of fixing their meaning of representation to avoid the claim becoming either obviously false or unfalsifiable. We use the framework of Dienes and Perner (1999) to provide a well-specified possible version of the claim, in which all representations of a minimal degree of explicitness are postulated to be conscious.

Perruchet & Vinter (P&V) offer the claim that "all mental representations are conscious" as a useful assumption for psychologists to adopt in interpreting data from learning, perception, language, memory, reasoning, and other areas in psychology. We will consider further this central claim of P&V's article. The truth of their claim, naturally, depends on what they mean by a representation. They admit that on a common understanding of the term (e.g., a pattern of activation coding an entity or class of entities in the world) "most representations are fully unconscious" (sect. 1.3.2). So what do P&V mean by representation, that makes their claim plausible?

Throughout their paper, "the word 'representation' designates a mental event that assumes the function of some meaningful component of the represented world (e.g., a person, an object, a movement, a scene) within the representing world" (sect. 1.3.2). This definition does not make clear whether "representation" is supposed to designate *representational vehicles* (e.g., neural states or processes that represent) or *representational content* (that which is represented).

The expression "mental event" strongly suggests the latter reading. Unfortunately, three problems arise: (1) It cannot account for thoughts about nonexisting entities, for example, a *unicorn galloping.* (2) It uses "mental" in its *definition.* Hence, it doesn't give us what we need, namely, an explanation of the mental in terms of representations (Field 1978; Fodor 1978). (3) Because "representation" is defined in terms of "mental event" and since the only way specified in the text of how we know of the existence of mental events is "through direct and personal [i.e., conscious] evidence" (sect. 1.4.1), P&V's central claim that *all mental representations are conscious* becomes practically tautological.

If we now try the other, albeit unnatural, reading that "mental event" refers to *neural processes* that have mental (representational) content, then we face the problem of how to delineate those neural processes that have representational content from those that don't. P&V's definition contains two suggestions in their next sentence: "At least two functions can be envisaged (Dulany 1997). A representation may evoke other representations (the representation of a pencil may evoke the representation of a pencil box, an exercise book, and so on). It may also enter as an argument into deliberative mental episodes (the representation of a pencil may be involved in reasoning, inference, action planning, and other mental activities)" (sect. 1.3.2).

Let us look first in more detail at the function of a "representation evoking another representation." One obvious problem with it is that it is not a function that can typically be *assumed by some meaningful component of the represented world,* as the main part of the definition requires. For example, evoking is not a function pencils have with pencil cases in the real world, even though thoughts (mental representations) of a pencil may well tend to evoke thoughts about pencil cases. Another problem is that almost all neural activation patterns call forth other activation patterns, but neural activation patterns per se are just the things P&V wish to exclude as representations. Hence this functional criterion totally fails to achieve the objective of distinguishing neural processes with mental content from those without.

We have much sympathy for seeing a close link between consciousness and P&V's second criterial function of being able to "enter as an argument into deliberative mental episodes." Unfortunately, two defects disqualify it for present purposes: (1) It is useless as a criterion for distinguishing neural processes with mental content from those without because it is no easier to distinguish neural processes that *enter as an argument into deliberative mental episodes* from those that don't than it is to distinguish neural processes with mental content from those lacking it. (2) The notion of *engaging in deliberative processes* is meaningful for us from our conscious experience. Hence using it as a criterion for neural processes being representations makes P&V's central claim that *all representations are conscious* tautological and unfalsifiable.

Without a clear notion of representation, any apparent case of unconscious knowledge in, for example, skill learning (Vinter & Perruchet 1999a) or any other domain can be dismissed as just based on mere activation patterns or a "processing pathway" rather than true mental representations per se. This renders the claim that *all mental representations are conscious* unfalsifiable.

One way of making sense of P&V's notion of representation and its relation to consciousness is to use the framework of Dienes and Perner (1999; 2002a; 2002b), who start with a notion of representation that is not defined in terms of the mental. They then develop a hierarchy of representations of different degrees of explicitness. This hierarchy offers P&V the opportunity to take only representations that are relatively explicit as being true mental representations (resolving a minor terminological difference). This would allow them to rule out the unwanted cases of activation patterns as cases of representation in a principled way and show why consciousness relates empirically to relatively explicit representations, that is, the "true representations" in P&V's sense that enter as arguments into deliberative mental episodes.

Dienes and Perner take functional theories (e.g., Dretske 1995; Millikan 1993) as the most promising accounts of a general notion of representation. Representations in this general sense include patterns of neural activation that have the function of indicating, for example, interaural time differences, cats, or any nonconceptual content. Connectionist weights are also representations whose functions are to indicate relations in the world. Clearly, this notion is too broad for P&V; it includes too much in the brain about which we are not conscious.

When we know a fact, for example, when we see that the word presented has the meaning *butter,* we have a certain propositional attitude (e.g., seeing) toward a particular proposition ("this word has the meaning *butter*"). In knowing the proposition, we form a representation (in the general sense) that explicitly represents some part or all of one's attitude and that state of affairs, resulting in a hierarchy of explicitness of knowledge.

At the bottom, (1) the representation might just make explicit a feature in the propositional content, for example, "butter." At the next level, (2) the feature is explicitly predicated of an individual, thereby explicitly representing the full proposition ("this word has the meaning *butter*"). Next, (3) whether the proposition is a fact or not is explicitly represented ("it is a fact that this word has the meaning *butter*"). Finally, (4) the propositional attitude by which the fact is known can be made explicit ("I see that [it is a fact that] this word has the meaning *butter*").

Why do we have the third stage of representing factuality explicitly? This may seem gratuitous. In fact, it is very important. Whenever a cognitive system can form different models of the same individual (Perner 1991) the system needs to be able to represent factuality explicitly. A child pretending a banana is a telephone is representing the same object as both a banana and a telephone (two models); but the child is not deluded, only one model is labeled as a fact. Factuality explicitness is also required for counterfactual reasoning and considering hypothetical situations as hypothetical. Explicit inference (using representations like, e.g., "If X then Y") requires representing that X is not actually the case, al-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

though there is an occurrent active representation about X. This can be contrasted with procedural knowledge. "If X then Y" can be implemented in a procedure in which a fact-implicit representation of X is connected to a fact-implicit representation of Y. In this case, any activation of "X" means X can be taken to be true, thus activating Y; that is, it implements the inference without explicitly representing it as an inference.

In sum, fact-explicitness enables inference, reasoning, planning, temporal understanding, and executive control (Perner 1998; in press), that is, deliberative episodes. It would be natural for P&V to use fact-explicitness for their definition of "representation." In fact that is how Piaget (1936/1953) seemed to use the term. All the abilities that require fact-explicitness tend to develop during the second year of life (e.g., Perner & Dienes, in press) at the end of the sensori-motor period that culminates in the acquisition of mental representation according to Piaget. This use of representation also conforms to Gibson's (1950) distinction between direct perception (transformations of completely implicit representations) and true inferences (that represent that whenever X is the case then Y must also be the case).

Consequently, P&V's claim would be that all occurrent fact-explicit representations in people are conscious. This is not an empty claim. In fact, Dienes and Perner (1999) argued that an additional level of explicitness is required for conscious awareness. In line with higher-order thought theory of consciousness (Carruthers 2000; Rosenthal 1986; 2000), it is necessary to also represent explicitly the mental attitude one has toward the fact in question. Thus, P&V's position would amount to the opposing claim that if a representation is fact-explicit, it will automatically be attitude-explicit (and hence conscious) as well. In this way, the central claim of their paper – "all mental representations are conscious" – ceases to be vacuous. (See Perner & Dienes, in press, for further discussion on the relation between fact-explicitness and attitude-explicitness; and Dienes & Perner 2002a; 2002b, for examples that show the atypical case of fact-explicit representations that do not become attitude-explicit, e.g., hypnosis).

## Mentalistic metatheory and strategies

Donelson E. Dulany

*Department of Psychology, University of Illinois, Champaign, IL 61821.*
**ddulany@s.psych.uiuc.edu**

**Abstract:** Mentalism (Dulany 1991; 1997) provides a metatheoretical alternative to the dominant cognitive view. This commentary briefly outlines its main propositions and what I see as strategies for its use and support at this stage. These propositions represent conscious states as the sole carriers of symbolic representations, and mental episodes as consisting exclusively of conscious states interrelated by nonconscious operations.

The *mentalistic metatheory* is a set of general propositions that provide a framework for more specific theoretical formulations, including those Perruchet & Vinter (P&V) use in so ingeniously reinterpreting an impressively wide range of phenomena. But I want this commentary to be complementary as well as complimentary, and so this outline is intended to support the present analysis and perhaps contribute to further analyses.

1. Mental episodes consist exclusively of conscious states interrelated by nonconscious mental operations, operations that are no more than the neural processes that interrelate those states.

2. Conscious states are the sole carriers of *symbolic* representations, which are functionally specified: They evoke other symbols, can be subjects and predicates of propositional thought, including the key proposition that "this represents that," and they participate in control of actions that give warrant to that key proposition.

3. First-order awareness symbolically represents something outside awareness. Higher-order (reflective) awareness symbolically represents a first-order awareness or mental episode by remembrance and/or inference.

4. Conscious states consist of various familiar modes – belief, perception, wish intention, and so on – and these modes carry contents.

5. Conscious contents can be propositional (a belief or perception that ___) or sub-propositional (a sense of ___).

6. In deliberative mental episodes, propositional contents are interrelated by deliberative operations of inference and decision. In evocative mental episodes, sub-propositional contents are interrelated by associative-activational operations – as in the authors' analyses. The episodes are interrelated: Through repetition, "B is predicated of A" may be become "A activates B." And "A activates B" may be represented in higher-order awareness as "B was produced by A."

7. Conscious contents come in "identity codes," in which things are identified as such, and "literal codes," which represent intensities, forms, colors, and the like. Literal codes precede (pre-attentively) and surround (as fringe) identity codes.

8. The domain of mental episodes runs from the output of sensory transducers through input to motor transducers.

9. Inactive memory consists of nonsymbolic neural networks, established through associative or deliberative learning (not "stored"), and they generate conscious contents when activated (a process that is not "retrieval").

10. Because all this elaborates the function of symbolic representations, it provides a modern analysis of what Brentano (1874/1973) called the intentionality of consciousness.

One *strategy* is emphasized by these authors (4, 5, 6). They show convincingly that associative-activational episodes interrelating conscious contents are *sufficient* to explain learning in a number of phenomena that have been attributed to complex unconscious processes: word extraction, word-object mapping, various syntactical patterns, as well as relational coding in transfer across common patterns with different content. Although PARSER, like any model, is neither conscious nor unconscious, it does what models can do – refine and extract the implications of theoretical assertions, which in this case consist of nonconscious associative-activational relations among conscious contents exclusively.

The authors emphasize the greater parsimony of their account, and that is valuable. But because these are competing process claims – theoretical formulations with constructs realistically interpreted – Bayesian relative credibility is the more significant test. The authors' interpretations of the data gain credibility not simply from parsimony but from the rich range of phenomena successfully interpreted without implausible auxiliary assumptions.

A second strategy can be simply expressed by saying that we see hypotheses of unconscious symbolism progressively disconfirmed when positive evidence consists only of weak and fragile effects that are methodologically questionable (target article, sect. 8.2) This is the nature of the authors' critique of unconscious perception and unconscious learning in normals. For example, relative credibility can be invoked by asking if evolution would have provided us with a mechanism for unconsciously perceiving reality, but it reveals itself only weakly in experiments that have been unreplicated or methodologically questionable.

With use of a little more of the metatheory, a more elaborate strategy becomes possible: We can examine the relative credibility of the standard claim that various neuropsychological syndromes provide a dissociation between consciousness and unconscious processes versus the mentalistic view that these syndromes produce what are perhaps more interesting dissociations *within* consciousness (Dulany 2000). Let's consider three the authors discuss (sects. 8.3, 8.1).

"Blindsight," the Oxonian oxymoron, may be more credibly interpreted as a dissociation between first-order consciousness of the stimulus presented and a higher-order conscious belief that this is "seeing" (Dulany 2001) – a dissociation following from damage to V1 (striate cortex) that impairs vision. The standard assess-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

ment said to reveal unconscious perception, commonly discrimination of one movement direction versus another (Weiscrantz 1997), is simply the well-known "direct objective measure of awareness" long recognized as the most sensitive available (Reingold & Merikle 1988). However low the confidence in his degraded awareness, the subject reports on every trial a conscious belief that the stimulus is more likely one thing than the other. Furthermore, for the now standard measure of awareness with the "commentary key," the subject "was instructed that he was to press the 'yes' key whenever he had any experience whatsoever of the visual event" (Weiskrantz 1997, p. 64) – clearly a report of higher-order awareness of first-order seeing.

Similarly, prosopagnosia may be more credibly interpreted, not as "unconscious recognition by an intact recognition system," but as a dissociation between literal awareness of facial form and identity awareness of the face – a dissociation produced by lesions around the fusiform gyrus. Put simply, we normally recognize faces from how they look. With Farah (1994) we can agree that these patients have damaged recognition systems, with their systematic performances controlled by residual capabilities – more specifically, I would add, by literal awareness of facial form and what that can activate. Relative credibility? These patients provide clear evidence of literal awareness of facial form in their successful matching of facial photos, whether familiar or unfamiliar (DeHaan et al. 1987). The lesions are also in the region in which activation passes from early visual processing to object identification. Furthermore, there is abundant evidence of neural activity in various loci associated with various conscious modes – not some separate system in the brain that might be "disconnected" from an intact recognition system.

We can also re-examine the claim for a dissociation between "conscious memory" and "unconscious memory" in anterograde amnesia (and also with well-known manipulations for normals). Explicit memory involves nonconscious deliberative operations on propositional contents. Implicit memory is evocative memory, with nonconscious activational operations on sub-propositional conscious contents. It is enough here to take the example of repetition priming: Presentation of the prime should strengthen the activational association of literal awareness of form, the graphemic level, to identity awareness of the word, the lexical level – as revealed in identification of the target word under impoverished stimulus conditions. The prime should also strengthen activational associations sequentially within the word, as revealed in more successful fragment completion of the target. What is presented to awareness becomes more available to awareness, with no evidence that what is not consciously remembered – "that the prime was presented" – is unconsciously remembered. With selective impairment, the dissociation is between kinds of mental episodes, each with interrelated conscious contents.

There is also a strategy of experimentally examining a rich set of interrelations among private reports of propositional awareness (e.g., Carlson & Dulany 1988), a strategy most useful with deliberative processing, but also useful where activational relations may be reported as propositions in higher-order awareness (e.g., Dulany et al. 1984). Where conditions for report validity are more questionable, the authors appropriately examine the rich set of relations among states of sub-propositional awareness clearly produced by the stimuli presented.

Eliminativists who endorse the computational view of mind, the view that consciousness is only a sometime and nonobligatory emergent of cognitive computations, have no satisfactory rationale for the evolutionary value of consciousness. The same is true of the many who endorse the standard information processing model in which complex mental episodes may occur in working memory outside conscious attention. Mentalism provides that rationale with consciousness as the sole carrier of symbolic representations of the not-here and not-now – the present in perception, the past in remembrance, and the future in expectations, wishes, intentions. Both standard traditions embody a view that even Freud worried about – in passing: that postulation of an unconscious might be no more than expression of the "primitive animism that caused us to see copies of our own consciousness all around us" (Freud 1915/1957, p. 171). A mentalistic metatheory leaves no place for a resident humanoid that secretly does what persons do – the cognitive unconscious.

## The limited roles of unconscious computation and representation in self-organizational theories of mind

Ralph D. Ellis

*Clark Atlanta University, Atlanta, GA 30314.* **ralphellis@mindspring.com**

**Abstract:** In addressing the shortcomings of computationalism, we should not throw the baby out with the bathwater. That consciousness is not merely an epiphenomenon with optional access to unconscious computations does not imply that unconscious computations, in the limited domain where they do occur (e.g., occipital transformations of visual data), cannot be reformulated in a way consistent with a self-organizational view.

For half a century, the notion of unconscious "computations," of which conscious states are epiphenomena, has been so entrenched that it is refreshing to see Perruchet & Vinter (P&V) present a case for the extreme opposite position. Even those who stop short of their extreme alternative conclusion should seriously question the assumption that unconscious computations underlie all conscious processes, with consciousness an epiphenomenon floating over this more basic mechanism.

Many commentators may charge overgeneralization beyond the data. No doubt, the stark, extreme viewpoints presented are not the only alternatives; the truth may fall somewhere short of the thesis that there are no unconscious computations in the brain whatsoever. But even the moderate view, that *many* conscious processes do not merely reflect unconscious computation, presents interesting questions about the fundamental meaning of "computation," "representation," and "mentality."

Part of the dilemma of computationalism stems from features of consciousness owing to its self-organizational nature: seriality, slowness, gradual decay, limited capacity, and embeddedness in valenced action – features reminiscent of Pribram's (1971; 1991) "holographic" theory of imagery. Gradual decay, seriality, and the like, are the very facts that make computationalist theories of imagery problematic (Kosslyn 1983; Paivio 1986; Poggio & Koch 1985). Similar points can be seen for gradual memory decay, fuzzy categories, fuzzy reasoning, and so forth (Ellis 1995). Pribram's theory of imagery, with the brain transforming inputs holistically rather than piecemeal, treated pattern and structure, not specific micro-components, as correlates of consciousness – a short step from the P&V self-organizational view as an alternative to computationalism. But does this entail that all mentality is conscious?

On the P&V characterization of standard theories of unconscious computations, we compute on received inputs, whose isomorphic imprints in the nervous system function as unconscious "representations" – hence the huge body of literature in analytic philosophy attempting to explain how isomorphic patterns in the nervous system function as "representations" of objects without making reference to any consciousness of the "aboutness" relation so that there can be unconscious representations (Bickhard 1993; Dennett 1987; Fodor 1981; Levine & Bickhard 1999; Pylyshyn 1973). But without making reference to consciousness (or to the subject's use of the object in action), only isomorphism and causal dependence connect the object and its supposedly mental representation. Neither isomorphism nor causation, of course, can make one thing represent another. As Thelen et al. (2001) point out, any dynamic internal event causally related to something external would then constitute a representation. In response, Newton (2002) insists that representation does occur – through a subject's *use* of it *in context* of embodied (organismically purposeful)



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

action. This formulation is congenial to the P&V project, but is less extreme, because it allows for unconscious representation: R represents object O if, because of some appropriate isomorphism, R plays the role of O in a simulated action involving O, either consciously or preconsciously (e.g., habituatedly). Given this possible formulation, the P&V exclusion of nonconscious representation seems unnecessarily counterintuitive.

In standard computational views, unconscious representations undergo transformations, resulting in behavioral outputs. These transformations of the unconscious representations are what constitute the unconscious computations. Consciousness is merely an optional way to access the results of the computations. For many computationalists, this renders consciousness an epiphenomenon, whose only causal powers over behavior or thought would have to be illusory (Jackendoff 1987).

Even those convinced that consciousness is not merely the epiphenomenon of information processing, but also requires processing in the unique manner of an active, self-organizational system, should notice that P&V force a choice between extreme viewpoints and ignore much middle ground. One successful research program frequently touted as perfectly compatible with computationalism involves different layers of sensory cortex in occipital and temporal lobes performing computations on incoming perceptual signals (Hubel & Wiesel 1959; Richardson 1991). These transformations are obviously unconscious, because when there is virtually complete occipital activation in response to a completely unexpected stimulus (indicating that the transformations are virtually complete), the subject still lacks perceptual consciousness unless there also occurs a parietal 300P electrical potential (Aurell 1983; 1984; 1989; McHugh & Bahill 1985; Srebro 1985; Weiskrantz 1986). Occipital and temporal lobes can do everything they normally do in processing the perceptual data, including the 100P occipital potential and the 200N "mismatch negativity," without the subject having consciousness of the stimulus. (In ERPs, the numbers refer to milliseconds after presentation of the stimulus. Extensive processing occurs during the first 250 msec of processing, with or without the consciousness accompanying the 300P.) These unconscious occipital transformations fit a computational paradigm: Cells in consecutive layers of sensory cortices analyze different features of perceived objects – lines, angles, shapes, colors, and so forth. These sequences of transformations are unimaginable on a conscious basis; we cannot imagine consciousness of color without shape or vice versa, yet our sensory cortices "compute" these properties separately and then recombine them.

Notice that the occipital transformations of perceptual signals are used to explain how a certain type of representation *comes about* in the first place. In one sense, we think of the pre-occipital signal (as received in the thalamus, for example) as an "unconscious representation," which will then be combined with other signals and transformed into a more fully developed "representation" – a representation in a different sense. But this highlights the need to think more carefully about what constitutes a "representation": Newton's enactive, self-organizational view of representation would eliminate the problems just mentioned by treating occipital activities as "potential" representations – activities that will lend themselves to use by the organism in representational action if the occasion should arise. The same could be said for representations in memory, thought, unconscious emotions, and the like. Representations and computations can occur without all conscious processes reducing to them.

## *Natura non facit saltum*: The need for the full continuum of mental representations

Robert M. French

*Department of Cognitive Science, University of Liège, Liège, Belgium.*
**rfrench@ulg.ac.be**

**Abstract:** *Natura non facit saltum* (Nature does not make leaps) was the lovely aphorism on which Darwin based his work on evolution. It applies as much to the formation of mental representations as to the formation of species, and therein lies our major disagreement with the SOC model proposed by Perruchet & Vinter.

Perruchet & Vinter (P&V) admit, of course, that conscious representations emerge from an underlying neural substrate. But the type of emergence for which they argue seems to involve a sudden, quantal leap from the unconscious to the conscious. One moment, the representation of an object, a scene, or a situation is in the process of being generated and is of no importance whatsoever in any cognitive calculus; and then, suddenly, as if by magic, the representation bursts into consciousness, thereby becoming endowed with all the cognitive powers of conscious representations. P&V write:

> mental life comprises only two categories of events: The conscious representations and the unconscious processes generating those representations. The two are linked like the head and tail of a coin. . . [the] processes and mechanisms responsible for the elaboration of knowledge are intrinsically unconscious, and the resulting mental representations and knowledge are intrinsically conscious. (target article, sect. 1.3.1)

It strikes us that a gradualist picture of representation-formation – for example, the classical Hebbian cell-assembly framework will suffice – would, in one fell swoop, explain most, if not all, of the instances of unconscious influences on conscious processing that the authors work so hard to explain away within their no-unconscious-representations SOC framework. Further, this new framework would in no way undermine the associationist principles that drive their model (correctly, in our opinion). Accepting the existence of representations that run the gamut from the embryonic and unconscious to the fully formed and conscious in no way implies the need for a "sophisticated unconscious processor."

The authors accept the notion of graded and partial conscious representations. Within a simple neural network framework, there is no problem extending these notions to unconscious representations. As it stands, the authors would have a great deal of difficulty in their SOC framework in distinguishing between an "unconscious representation" and an "absent representation" (i.e., no representation), because they would maintain that both situations have no effect whatsoever on conscious perception. But, as we hope to show in the thought-experiment presented here, there must be a difference. This difference, if a real-world version of the thought-experiment were actually run, would presumably be able to be measured with appropriately sensitive instruments.

Suppose that two individuals, A and B, start with perfectly identical brains. Via a rigid, completely reproducible procedure, A learns the concept $\Omega$, and B does not. Now, presumably, learning $\Omega$ involves a physical (presumably, synaptic) modification of a specific set, $S_A$, of neurons in A's brain. The precisely corresponding set of neurons in B's brain, $S_B$, undergoes no such physical change. Presumably, P&V would say that the concept $\Omega$ is now physically represented in A's brain (whether active or not). Now, since they explicitly accept the concept of representational decay, we will suppose that the synaptic changes that constituted A's representation of $\Omega$ gradually decay in precisely the reverse order in which they were strengthened when A was originally learning the concept $\Omega$. Further suppose that we have a device capable of stimulating the neurons in $S_A$ (and only those neurons). At some point during this decay toward the original state of the neurons before A learned $\Omega$, A would presumably no longer be consciously aware



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

of the concept Ω when $S_A$ was stimulated. (This point will be somewhere in the zone corresponding to A's very early learning of the concept, before the representation would be conscious. P&V explicitly concede that there is such a period.) At this point, we now have A relearn Ω and B learn Ω for the first time, employing exactly the same procedure originally used when A first learned Ω. Surely, P&V would agree that A would relearn the concept Ω faster than B because, as we have set things up, A will have a representational "head-start" over B. We thus have a very simple hypothetical case of how an unconscious representation could significantly affect the conscious experience of concept acquisition. Further, the SOC account, relying as it does only on conscious representations, would be at a loss in explaining this learning-time difference, unless they took the unfalsifiable position that A's more rapid learning of Ω simply demonstrated that the decayed representation with which A started prior to relearning Ω must, in fact, have been conscious all along.

It may well be that there is, indeed, some sort of "connectivity phase change" when a neural representation has the possibility of becoming conscious when activated. This could be the point described by Hebb as when "reverberation in the structure might be possible . . . reverberation which might frequently last for periods of time as great as half a second or a second, [this being] the best estimate one can make of the duration of a single 'conscious content'" (Hebb 1949, p. 74). But if one is to present a coherent picture of cognition that takes into account neural, representational, and cognitive phenomena, one must not neglect the representational stages leading up to this creation of cell-assemblies or, in the language of P&V, up to the emergence of fully conscious representations.

In conclusion, we suggest that the SOC model might do well to turn to basic neural network principles that would allow it, without difficulty, to encompass unconscious representations, as described above. (See, e.g., Cleeremans & Jiménez 2002; Mathis & Mozer 1996.) These "unconscious" representations – some of which may evolve into representations that, when activated, would be conscious – can affect consciousness processing, but do so via the same basic associative, excitatory, and inhibitory mechanisms that we observe in conscious representations. The inclusion of this type of representation in no way requires the authors to also posit sophisticated unconscious computational mechanisms.

## Unconscious semantic access: A case against a hyperpowerful unconscious

Daniel Holender[a] and Katia Duscherer[b]

[a]*Laboratoire de Psychologie Expérimentale, Université Libre de Bruxelles 50, Brussels, Belgium;* [b]*Laboratoire de Psychologie Expérimentale, Université René Descartes (Paris V), Boulogne-Billancourt, France.*
**holender@ulb.ac.be** **kduscher@ulb.ac.be**

**Abstract:** We analyze some of the recent evidence for unconscious semantic access stemming from tasks that, although based on a priming procedure, generate semantic congruity effects because of response competition, not semantic priming effects. We argue that such effects cannot occur without at least some glimpses of awareness about the identity and the meaning of a significant proportion of the primes.

Like Perruchet & Vinter (P&V), we fully endorse a mentalistic perspective, which implies that we do not posit the existence of a "powerful," or more precisely, an *intentional* cognitive unconscious. Thus, we basically share the view of Searle (1990; 1992) and Dulany (1997) that the unconscious is intentional in a dispositional way. In this commentary, we expand on the claim made by P&V in section 8.2 that the available data on unconscious semantic access do not constitute a challenge to the mentalistic framework.

In assessing the plausibility of the evidence for unconscious semantic access, a distinction must be made between tasks generating semantic priming effects and tasks generating other effects based on stimulus meaning, such as Stroop and Stroop-like congruity effects. This distinction has been somewhat blurred in recent work, maybe partly because of the multiple meanings of the term *priming*, which can designate an experimental procedure, an observed effect, and a hypothetical causal process, such as automatic spreading activation in semantic memory (e.g., Neely 1991). Much of the early evidence for unconscious semantic access under masking, criticized by Holender (1986), was based on a semantic priming paradigm yielding bona fide semantic priming effects. Much of the recent evidence for unconscious semantic access discussed by P&V does not qualify as priming because it rests on tasks that, although based on a priming procedure, are functionally equivalent to Stroop-like tasks. These tasks are generally assumed to generate congruity effects because of response competition (e.g., Eriksen 1995; Holender 1992; MacLeod 1991), not priming effects.

The studies of Greenwald et al. (1996; Draine & Greenwald 1998) are based on prime and target words with strong positive and negative affective connotations. The SOA between the prime and the target is very brief (under 100 msec), and the prime is interleaved between two masks consisting of random letters strings. Even though the primes could not be discriminated above chance, the binary classification of the target words in terms of their pleasantness is more accurate in congruent trials, in which the polarity of the prime and the target words are the same, than in incongruent trials, in which the polarities are opposite. Similarly, in the studies of Dehaene et al. (1998; Naccache & Dehaene 2001), which are based on a comparable procedure, the speed of classification of a single-digit target number in terms of whether it is larger or smaller than five is affected by the congruency of the unconscious prime number.

Initially, Greenwald et al. (1996; Draine & Greenwald 1998) interpreted their finding as reflecting semantic priming based on spreading activation. Then, Klinger et al. (2000) demonstrated that this effect does not depend at all on spreading activation but on response competition. This was taken as evidence that the unconscious primes must be covertly classified according to the same rule as the one applied to the visible target (see also Dehaene et al. 1998). Next, it was shown that the congruity effect only appears with primes that have been used repeatedly as targets (Abrams & Greenwald 2000; Damian 2001), which prompted a reinterpretation of the effects in terms of the formation through learning of a direct stimulus-response link based on superficial features of the stimuli. However, Abrams et al. (2002) argued that this link must rather be established between the stimuli and the semantic categories, as the learning effect resisted a change in response assignment. Nevertheless, Naccache and Dehaene (2001) persisted in their account in terms of unconscious semantic classification, because the congruity effect still occurs with unconscious primes, which have not been seen before as targets.

All these interpretations of unconscious congruity effects rest on the assumption that the primes are completely unavailable to awareness. If correct, they imply a hyperpowerful unconscious, that is, an unconscious even more powerful than the one already required to explain unconscious semantic priming effects. We contend that this conception is profoundly mistaken because, as was pointed out by Prinz (1997), a stimulus has no inherent information sufficient to specify a response outside the context of a goal-directed task imposed by the instructions. Besides, the primary source of response conflicts underlying the congruity effects described above must lie in conscious mental representations (cf. Holender 1992), because there is no stored information, and hence no information that can be automatically activated, about whether a number is smaller or larger than five or about whether the concept denoted by a word has a pleasant or unpleasant connotation. Therefore, the only possible source of conflict lies in the fact that most participants think about the irrelevant information in terms similar to those used by the instructions to describe how


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

the relevant information has to be related to the responses. Actually, we have evidence stemming from a Stroop-like task with visible distractors that participants not noticing the critical relation between the irrelevant information and the responses do not show any congruity effect (Duscherer & Holender, submitted).

To conclude, we submit that none of the congruity effects discussed above could occur without at least some glimpses of awareness about the identity and the meaning of a significant proportion of the primes. However, not all the apparent evidence of unconscious semantic access is flawed. There are also replicable effects that, although puzzling at first sight, can now be elegantly accommodated within the mentalistic framework. Let us take three examples. First, there is no need to postulate a center-surround attentional process taking place in semantic memory, to account for the reversal in the semantic priming effect (faster responses in unrelated than related trials) when participants make a semantic similarity judgment on masked primes in a preliminary task (Carr & Dagenbach 1990; Dagenbach et al. 1989). Kahan (2000) accounts for this effect by a retrospective prime clarification process stemming from participants' deliberate attempt to extract the meaning of the masked word, with this strategy being itself promoted by the preliminary task. Second, Duscherer and Holender (2002) explain the unconscious negative priming effects found by Allport et al. (1985) by the fact that a distractor potentially available to awareness can be made unconscious by the synergy between an act of selective inhibition taking place during prime processing and attentional diversion elicited by the onset of a mask coming too late to cause much perceptual degradation. Third, it is not necessary to postulate a powerful unconscious that can make affective preference judgments on stimuli that cannot be recognized (Zajonc 1980). This effect can simply result from applying a more effective nonanalytic strategy in preference judgments, and a less effective analytic strategy in recognition judgments (Whittlesea & Price 2001).

ACKNOWLEDGMENT

Katia Duscherer was supported by a postdoctoral grant from the Fyssen Foundation.

## Consciousness and unconsciousness of logical reasoning errors in the human brain

Olivier Houdé

*Groupe d'Imagerie Neurofonctionnelle, UMR 6095, CNRS, CEA, Université de Caen (Centre Cyceron) and Université Paris 5, Institut Universitaire de France, Sorbonne, 75005 Paris, France.* **olivier.houde@paris5.sorbonne.fr**
**www.cyceron.fr**

**Abstract:** I challenge here the concept of SOC in regard to the question of the consciousness or unconsciousness of logical errors. My commentary offers support for the demonstration of how neuroimaging techniques might be used in the psychology of reasoning to test hypotheses about a potential hierarchy of levels of consciousness (and thus of partial unconsciousness) implemented in different brain networks.

Since Aristotle, we have known that the essence of the human mind is the *logos*, that is, both reason (logic) and language. But the seventeenth-century French philosopher Descartes (1628/1961) also showed with his method that an important challenge for humans is to implement deductive rules in order to redirect the mind from reasoning errors to logical thinking.

My commentary is specifically about section 7 of the target article, that is, about problem solving, decision making, and automaticity. Recent cognitive psychology and neuroimaging studies by my group have dealt with the mechanisms by means of which the human brain corrects initially unconscious logical reasoning errors (Houdé et al. 2000; 2001). They show that the activated brain networks are different, depending on whether (1) subjects think they are responding correctly to a logic problem when in fact their reasoning is biased by an erroneous perceptual strategy (an automatic strategy, in accordance with Evans's (1989) model); or (2) they become aware of their error and correct it after being trained to inhibit the initial perceptual strategy. In the second stage (after training), regions in the left lateral prefrontal cortex devoted to executive functions, inner speech, and deductive logic are activated, along with a right ventromedial prefrontal area dedicated to self-feeling and relationships between emotions and reasoning (see Damasio's theory on consciousness; Damasio 1999).[1] None of these regions are implicated in the first stage (before training), where the only activation observed is in a posterior network strongly anchored in perception (ventral and dorsal pathways).

Interestingly, from the famous case of Phineas Gage in the nineteenth century (Damasio et al. 1994; Harlow 1848) to Damasio's patients today (Damasio 1994; 1999), neuropsychological findings clearly indicate that right ventromedial prefrontal damage is consistently associated with impairments of reasoning/decision making, emotion, and self-feeling. For the first time, our neuroimaging results demonstrate the direct involvement, in neurologically intact subjects, of a right ventromedial prefrontal area in the making of logical consciousness, that is, in what puts the mind on "the logical track," where it can implement the instruments of deduction. (Note that this brain area was not activated in a group of subjects who were unable to inhibit the initial perceptual strategy and therefore could not avoid the reasoning error; see Houdé et al. 2001.) Hence, the right ventromedial prefrontal cortex may be the emotional component of the brain's error correction device. More exactly, this area may correspond to the self-feeling device that detects the conditions under which logical reasoning errors are likely to occur (in connection with the anterior cingulate cortex; see Bush et al. 2000).

From the standpoint of evolutionist psychology (Bjorklund 1997; Tooby & Cosmides 2000), it is interesting to relate these neuroimaging results to the role classically ascribed to emotions in survival, namely, that in the face of danger (here, logical reasoning errors), fear leads animals and thus humans to flee, to avoid. In Darwinian terms, we can contend that evolution must have fashioned the brain to sense the emotions needed to inhibit nonadaptive behavior, even nonadaptive reasoning strategies (Houdé et al. 2000).

The findings of these studies (Houdé et al. 2000; 2001) allow me to challenge Perruchet and Vinter's concept of self-organizing consciousness (SOC) in regard to three interrelated points: (1) the existence – denied by the authors – of a hierarchy of levels of consciousness (and thus of partial unconsciousness) implemented in different brain networks (see Stage 1 and Stage 2 in our training paradigm below); (2) the important problem in adults, but also in children (Houdé 2000; Piaget 1984), of the consciousness or unconsciousness of reasoning errors (an issue the authors do not really address in the article); and (3) the question of the involvement of emotions and self-feelings in logical reasoning, if and when the subject becomes aware of the fact that there are several ways to solve a problem (and that one of them is based on a misleading strategy).

NOTE

**1.** *The Feeling of What Happens: Body and Emotions in the Making of Consciousness.* I am surprised to find no references to Damasio's work, and in particular, with regard to the issue of decision making, no mention of the 1997 article in *Science* that he co-wrote (Bechara et al. 1997).



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

## Surfing on consciousness, or, a deliberately shallow outline of cognition

Luis Jiménez

*Department of Psychology, University of Santiago, Campus Sur, Santiago, 15782, Spain.* **jimenez@usc.es** **http://web.usc.es/~psljim/index.html**

**Abstract:** By assuming that conscious states are the only constructs entitled to bear a cognitive status, while denying this status both to the learning processes and to their nonconscious outcomes, the SOC view leaves consciousness alone as the single tool to explain itself. This does not endow consciousness with any self-organizing properties, but rather, draws a deliberately shallow outline of cognition.

Going through this target article, I found myself in agreement with a good number of its major proposals, and yet my overall reaction to the SOC view was one of disappointment. On a closer look, I realize that this reaction was probably because of the fact that, by assuming that nothing cognitive can happen below the threshold of consciousness, this view of cognition has taken its lack of depth as axiomatic. In this commentary, I would like to suggest that, quite on the contrary, the cross-talk between different levels of description is essential for a cognitive account to convey any sense of explanatory depth (Cleeremans & French 1996), and that this explanatory power is much more easily obtained if we undertake a cognitive analysis on both the unconscious processes of learning and the unconscious knowledge structures that support the emergence of conscious states.

To put these arguments into a wider context, I will start by stressing the commonalities existing between the SOC view and the position that I defend in this commentary (see Cleeremans & Jiménez 2002, for a principled presentation of this view). The three main points of the SOC view with which this position would agree have to do, respectively, with the challenge to the computational view of mind, the vindication of a central role for conscious states, and the dynamic relation postulated between learning and consciousness. More specifically, I agree that the computational theory of mind, with its view of cognition as the mere shuffling of formal symbols (Searle 1999), cannot be considered as a serious candidate to account for the workings of the conscious mind. Second, I also share the conviction, expressed by the proponents of the SOC view, that conscious representations should play an important role in the dynamics of cognition, and that they are essential not only to govern the current behavior of the cognitive systems, but also to determine their future adaptations. Third, I wholeheartedly agree with the claim that consciousness and learning are intimately related, in that one of the main functions of learning is to change consciousness in an adaptive way, whereas one of the main functions of consciousness is to poise the most relevant representations for further learning.

However, where my position starts to depart sharply from that of the SOC view is when it comes to characterizing the nature of the learning processes, and of the unconscious structures that must hold the knowledge of the system. The SOC view describes learning as a set of nonconscious, elementary, and associative processes operating on conscious representations, but it denies its cognitive status; and hence it skips a detailed, functional analysis of its workings. Moreover, it explicitly denies the existence of any nonconscious structures that might play any role in storing and transforming the knowledge of the system while this knowledge is not phenomenally enacted. In so doing, it strips cognition from any explanatory burden and describes it as a shallow structure that merely appears to *surf* on the waves of consciousness. However, despite the authors' efforts to draw this shallow outline of cognition, the murmur of deep waters keeps sounding at some places, such as when they refer to internal units that are subjected to processes of consolidation and decay, or when they use PARSER as a model to simulate these processes.

In these and some other cases, the authors account for the dynamics of consciousness by appealing to structures that contain knowledge, and that are subjected to certain transformations even while they remain below the surface of consciousness. The SOC view still refuses to ascribe a cognitive status to these structures and to the processes that act on them, by arguing that such unconscious representations may be recast as mere dispositions (see Appendix A) and that the underlying processes are "purely physiological phenomena" (cf. Perruchet & Gallego 1997).

However, it is not clear to me how these arguments could be any more convincing when they are leveled at the existence of nonconscious representations and processes than they were many decades ago, when similar concerns were raised by philosophical behaviorism against the explanatory relevance of any mental events. Indeed, the concepts of mental representation and processing resisted these attacks, not because they were not enunciable in dispositional terms or because they could not be ultimately reduced to physiological phenomena, but rather because they provided far better primitives within which to undertake a functional analysis of cognition. In a similar vein, I submit that the computational analysis of unconscious cognition should also be maintained within a maturing view of cognitive psychology, because it provides us with essential, low-level tools from which to explain the functional underpinnings of consciousness.

Of course, this is not to say that the processes and representations underlying this unconscious cognition must be viewed as identical to those underlying their conscious counterparts. Hence, the authors are quite right in taking issue with the symbolic views that have attempted to cast unconscious processing as a rule-driven, combinatorial process performed by some intentional, but unconscious, agents. However, the concept of cognitive representation does not require the presence of such intentional interpreters, but may arise from the action of any system that learns, provided that it is wired properly to maintain the representing patterns in phase with their relevant meanings (Dretske 1988). With this less restrictive notion of cognitive representation in hand, cognition could exist even if these intentional but unconscious agents were completely banished. Indeed, even if we were able, at some point, to banish the homunculus in charge of the conscious control processes, so that no cognitive operation would need to be attributed to the action of such a full-blown intentional agent, the concepts of cognition would not vanish automatically with it; rather they would still provide us with an informative, functional story about how the mind works – precisely, about how the deep waters of our brains and minds can manage to bring about the brilliant surfaces of consciousness. The SOC view just invites us to surf on this marvelous surface, but I would say, no thanks, I prefer diving.

ACKNOWLEDGMENTS

This work has been supported by grant XUGA 21106B98 from the Xunta de Galicia (Spain) and by the DGES Grant PB97-0525 from the Ministerio de Educación y Cultura (Spain).

## Unconscious abstraction in motor learning

Aysha S. Keisler and Daniel T. Willingham

*Department of Psychology, University of Virginia, Charlottesville, VA 22904.*
**ask7x@virginia.edu** **willingham@virginia.edu**

**Abstract:** Given the complexity and generalizability of motor skills, it is difficult to account for learning in this area without incorporating the concept of unconscious abstraction. A model based solely on association does not seem to account for data in this domain; specifically, instances that require learners to execute a practiced motor skill in a novel situation.

Perruchet & Vinter (P&V) offer a strong logical case for the claim that all types of learning do not necessitate abstract rules, including tasks which do indeed follow such rules. Rather than learn an abstract underlying rule, they propose, the learner may instead



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

unconsciously associate elements of the stimuli that frequently appear together. The authors do not specifically discuss motor learning in their review of learning systems and the SOC framework, other than to point out that this domain is no exception in their association-based, abstraction-free model. However, a model that does not allow for abstract learning is difficult to reconcile with motor research, given the generalizability of many motor skills. Here, we discuss instances of abstract rule learning that do not appear reconcilable with P&V's associative hypothesis, specifically, motor learning in which new behaviors may be extrapolated from previous experience in ways inconsistent with association. The fact that humans can adapt complex motor skills to new situations offers compelling evidence of abstract and unconscious learning in the motor domain.

First, we agree with P&V that a task may follow a physical rule but the motor system need not use that rule. For example, McLeod and Dienes (1993) found that catchers ran to catch a ball so that $d^2(\tan a)/dt^2 = 0$, where $a$ is the angle of elevation from the fielder's gaze to the ball. However, catchers may also exploit a visual strategy so that they arrive at the point at which the ball will fall (Dienes & McLeod 1993; 1996; Michaels & Oudejans 1992), without the use of the equation.

In some tasks, however, unconscious computation appears necessary to successfully account for behavior. Shadmehr and Mussa-Ivaldi (1994) hold that, to compensate for an external force field during a motor task, the motor system creates an "internal model" of the field independent of position. In this study, subjects made reaching movements via a handle on a robotic manipulator in the presence of an external force field. Counterforce from the manipulator served as the force field. Over time, subjects learned to compensate for the external force and could perform as they did in trials with no force (null field). Shadmehr and Mussa-Ivaldi ruled out the possibility that subjects learned an association of external force and the necessary counterforce by observing the presence of after-effects when the null field was introduced after adaptation to the force field. Because there was no external force on these trials, the motor system should not associate the movement with a counterforce. They also ruled out the hypothesis that subjects learned via associations between the force and visited locations by observing that learners could compensate for the force field at unvisited locations. Furthermore, Conditt et al. (1997) reported that subjects exposed to a similar mechanical force field could create smooth movements different from movements practiced during training. This observation contradicts the possibility that subjects learn via associations between the movement trajectory or position and the appropriate counterforce, or "rote learning" in Conditt et al.'s terminology. Rather, the motor system computes the appropriate counterforce from knowledge of (1) the external force and (2) properties such as velocity, position, orientation, and so forth, of the hand. The fact that the system must *extrapolate* the appropriate counterforce based on this information is the critical evidence of unconscious "high-level conceptual processing," against which P&V argue.

Goodbody and Wolpert (1998) discuss velocity generalization in force-field motor tasks. These authors taught subjects a simple reaching movement while exposed to a velocity-dependent force field until subjects could successfully compensate for the external force. Then, subjects were asked to perform new movements and their ability to transfer knowledge of the field was observed. Goodbody and Wolpert concluded that "the force-velocity relationship is internalized in a functional form and then linearly extrapolated to new speeds" (Goodbody & Wolpert, p. 1825). The authors ruled out association-based accounts (including accounts allowing for generalization of the sort favored by P&V) because the generalization gradients did not conform to such a theory.

The studies mentioned here contain evidence of varying levels of learning transfer that indicate abstract, unconscious learning. Research by Felice Bedford offers similar evidence in the perceptual domain. In a series of prism-goggle experiments, subjects generalized visual-spatial mapping (Bedford 1989; 1993), indicating that subjects did not learn an association between a region of space and a visual-spatial mapping, but rather combined trials to form a global mapping.

P&V stress the lack of transfer evident in various domains and site this as evidence against unconscious abstraction. For example, the authors cite the various serial reaction time tasks in which implicit knowledge of a visuo-motor sequence did not transfer if certain features were changed. However, the *quantity* of cases in which learning does not transfer to new situations should not be interpreted as evidence against the existence of abstract learning. Under other conditions, learning transfers well (Willingham 1999; Willingham et al. 2000). In addition, subject performances under novel conditions in the force-field studies by Shadmehr and Mussa-Ivaldi and Conditt et al. were nearly identical to performance in a well-practiced field (Conditt et al. 1997; Shadmehr & Mussa-Ivaldi 1994). The pattern of transfer successes and failures can be informative regarding the representation of learning. In the instances we have cited, they indicate unconscious, abstract representations of motor learning.

## The conscious and the unconscious: A package deal

Martin Kurthen

*Department of Epileptology, University of Bonn, D-53105 Bonn, Germany.*
**martin.kurthen@ukb.uni-bonn.de** **http://www.epileptologie-bonn.de**

**Abstract:** Parsimony and simplicity in cognition theory are not achieved by excluding either the "cognitive unconscious" or consciousness from theoretical modeling, but rather, by eliminating redundant constructs independent of their location on the conscious-unconscious axis. Hence, Perruchet & Vinter's (P&V's) case against the "cognitive unconscious" does not work as an argument *for* consciousness, but rather as a rejection of the redundant background computational processes postulated in traditional cognition theory.

Perruchet & Vinter (P&V) cogently apply Occam's razor to the cognitive unconscious, but in my view they are misled in trying to establish the usual victim of that razor, namely, consciousness, as an alternative. In their model of self-organizing consciousness (SOC), the role of the cognitive unconscious is taken over by what they call "unconscious processes," which are brain processes on higher levels of organization. P&V show that these processes need not be conceptualized as inferences or unconscious representations according to their own idiosyncratic definition (sect. 1.3.2), but there is no positive argument to the effect that cognitive functioning requires a superimposed level of phenomenally conscious representations. The "cognitive conscious" may still be irrelevant to cognitive success, even if the cognitive unconscious has been dismissed. P&V's nonrepresentational "unconscious processes" could be linked to behavior without the consciousness interface.

This is obvious from P&V's first illustration of their critique (sect. 1.3.3). In the classical conditioning example that P&V use to illustrate the redundancy of unconscious inferences, it is associative mechanisms, not conscious representations, that take the place of unconscious inferences. To describe how S1 will achieve a "negative valence" by means of some learning procedure implemented in trivially "unconscious" brain processes, there is no need to refer to the accompanying perceptual experience. The message of this example is that classical conditioning can be understood without the recourse to additional mnestic and inferential processes. The crucial message is *not* that classical conditioning can be understood without recourse to any sort of *unconscious* processing. And the message *cannot* be that classical conditioning cannot be understood without the recourse to *conscious* representations. In fact, there is no necessity to have "associative principles" (sect. 9.1) operate on "conscious contents." These principles might just operate on a certain class of trivially "unconscious" brain



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

processes (perhaps *accompanied by* conscious episodes). Furthermore, self-organization as defined by P&V can be described in terms of content rather than consciousness (sect. 1.4.2). Hence, associative processes combined with the principle of self-organization together replace unconscious rule-governed thinking: It is not necessary to refer to an additional "consciousness." At least, P&V themselves do not argue for this necessity. They do not even hold that the "prevalent framework" is "unwarranted" (sect. 1.4). Their more moderate purpose is to sketch an alternative to the traditional framework of cognition qua unconscious representations, a framework in which the mental life is "co-extensive with consciousness" (sect. 1.3.1). In this framework, consciousness is taken as a sort of *via regia* to cognition, but why?

Looking for explicit reasons, one finds that P&V see starting with consciousness as the most natural and direct way of studying cognition. They invoke the notorious intuition that the "direct and personal evidence" (sect. 1.4.1) we all have of the "existence" of conscious mental states, somehow makes these states suitable for basic psychological modeling. In contrast, our unconscious mental life is but a postulate, something of which we have no "direct proof" (sect. 1.4.1). But that sort of "personal evidence" is not well qualified as a starting point for an empirical, scientific analysis of cognition. Aside from their highly problematic ontological status, phenomenally conscious states as such are methodologically unattractive for cognitive science (Kurthen et al. 1998). What makes them useful is their reportability (or other overt "markers" or correlates), by which they enter into meaningful behavior in scientifically utilizable settings. But then verbal reports (or other markers) as behavioral evidence, not conscious episodes as such, qualify as suitable for psychological modeling. Cognition theory can proceed with that behavioral evidence and leave the correlated conscious events aside.

Another, more implicit reason for P&V's insistence on consciousness could be that unexpressedly, they still aspire towards a *theory* of consciousness. In that case, the whole thesis would have to be reversed to say that, then, *the SOC assumption is the very thesis that allows one to explain phenomenal consciousness* (and the isostructurality between consciousness and the world). The SOC thesis would then be the entry to an explanation of consciousness, instead of consciousness being the starting point of an explanation of cognition. One would then hold that the very interplay of associative mechanisms and self-organization leads – among other things – to the phenomenally conscious episodes we have "direct evidence" of (episodes which might still be cognitively irrelevant). This would be an interesting path to follow, but it is not what P&V strive for, at least in my reading.

Why not take the opportunity to dispense with the cognitive unconscious *and* the "cognitive conscious" in psychological modeling by abandoning the *whole question* of whether or not a property or process or entity is conscious or unconscious? This is not meant as an attempt to eliminate or explain away consciousness (and the unconscious), but rather as a methodological rejection of the conscious – unconscious dichotomy as irrelevant to cognition theory. P&V's highly valuable proposal to *occam-ize* the (what they call) cognitive unconscious is artificially complicated by the recourse to that dichotomy. Cognition theory is not primarily about "the conscious" and "the unconscious": Orthogonally to the conscious-unconscious axis, it is about the whole organization of cognitive behavior. But once you project or map this organization onto that axis, you find that the conscious and the unconscious always come in a package deal.

If you privilege consciousness in psychological modeling, you are somehow condemned to care and find a place for the repudiated unconscious as well (and vice versa, as Freud had experienced a century ago). This dynamics reminds of the psychoanalytically described mechanism of the "return of the repressed," with the cognitive unconscious in the role of the repressed content in the mentalistic framework, annoyingly challenging the purity and unity of that framework out of its pen of unconscious nonrepresentational "processes." And the solution is: No repression, no unwanted return, all the more because the whole argument against the cognitive unconscious *as something unconscious* is a waste of theoretical energy. For the crucial property of the cognitive unconscious that P&V argue against, is not its not being conscious, but *its representational and inferential or, more generally, computational constitution independent of its location along the conscious-unconscious axis* (see the classical conditioning example discussed above). So why not realize the conscious *and* the unconscious as they came – low-priced, in a package deal? That would clear the field for an *occam-ization* of the entities P&V's argument really applies to: namely, that thicket of background processes of "reasoning, inference, and so on" (sect. 1.2), which the traditional information processing approach postulates for perhaps insufficient reasons.

## The reported demise of the cognitive unconscious is premature

Anthony J. Lambert

*Research Centre for Cognitive Neuroscience, Department of Psychology, University of Auckland, Auckland, New Zealand.* **a.lambert@auckland.ac.nz**

**Abstract:** Perruchet & Vinter (P&V) claim that all mental representations are conscious and that the cognitive unconscious does not exist. Unfortunately, support for these claims is garnered illicitly, by defining the terms *representation* and *computation* in a way that is unduly restrictive and misleading. In addition, their proposals fail to engage adequately with critically important evidence and theory from neuropsychology concerning consciousness and cognition.

Perruchet & Vinter's (P&V's) proposals include two provocative assertions: (1) That all mental representations are conscious, and (2) that the cognitive unconscious does not exist. The second may be viewed as an extension of the first, if cognitive processes are identified with computational operations operating upon mental representations. If there is no such thing as an unconscious mental representation, there is no such thing as unconscious computation (or unconscious cognition) either.

Unfortunately, P&V's analysis depends upon defining *representation, cognitive process,* and *computation* in a way that is restrictive and misleading. For P&V, the term *representation* "designates a mental event that assumes the function of some meaningful component of the represented world (e.g., a person, an object, a movement, a scene)" (sect. 1.3.2). They note that a pattern of neural activity does not necessarily constitute a representation; for example, the pattern of neural activity in the retina does *not* constitute a representation in their terms (sect. 1.3.2). However, it is clear that even at the level of the retinal ganglion cells, the information encoded by the rod and cone receptors has undergone a variety of sophisticated transformations. For most cognitive scientists, these transformations are not simply regularities that may be modelled in computational terms, like the weather or planetary movements – they *are* computations, which extract information such as luminance and colour transitions from the retinal image. These early (retinal ganglion) informational outputs then participate in a multitude of further neural transformations in the many cortical and subcortical brain areas that give rise to visual experience. P&V's analysis begs this question: At what point in processing does a "pattern of neural activity," which in their terms does not constitute a representation, cross a critical boundary and become a representation by assuming "the function of some meaningful component of the represented world"? (sect. 1.3.2; cf. Dennett & Kinsbourne 1992). P&V's answer is that this boundary is crossed when the resulting event can participate in the "evocative" or "deliberative" mental episodes proposed by Dulany (1997). This, together with the emphasis on *meaning* in their definition, suggests that semantic features, together with access to other semantic level representations, are critical. One reading of this



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

might be that patterns of neural activity in association cortex can constitute representations, while those in primary sensory (or motor) cortex do not. P&V's conception of representation is therefore restrictive, in referring only to relatively "high-level" representations, and explicitly excludes early "sensory-level" (pre-semantic) representations. Furthermore, P&V apply essentially the same restriction to the notion of a cognitive operation, viewed as computation. According to P&V, the term *computation* "does not extend to any neural activity, but instead designates the mental operations that take representations as arguments" (sect. 1.3.2).

These restrictive definitions of representation and of computation are indefensible. The notion of multi-level representation is intrinsic to *all* forms of computational architecture. For example, the laptop computer I am using to write this can represent the word *computation*, at the word level (when checking against entries in the spell checker), at the letter level, or at the level of the pixel mosaic on the display screen. Computations can be applied to any or all of these representations. Hence, P&V use the term *computation* idiosyncratically, in a way that no computer scientist would recognise. In addition, P&V use the term *representation* in a sense that is far narrower than that generally used in cognitive science. The problem here is that for several decades the area of cognitive science, where the use of an explicitly computational approach has, arguably, has enjoyed the greatest success, is the study of early vision (e.g., Marr 1982). However, according to P&V's analysis, computation does not happen during early (pre-semantic) visual processing!

P&V wish to defend the position that all representations are conscious, and all computations are conscious (the new mentalism). Unfortunately, this is achieved primarily by defining *representation* as meaning high-level (semantic) representation, and *computation* as meaning high-level (semantic) computation. This simply sidesteps a central task of cognitive science. In the case of both brains and computers, complex behaviour arises from the combined action of millions of individual components, each one of which is relatively simple (although individual neurons are undoubtedly far more complex than the individual components of a silicon chip). In the case of computers, complex computations and representations (such as those manipulated in object oriented programming) are constructed from the combined action of very simple representations and computations that manipulate those representations. The challenge facing cognitive science is to understand how mental phenomena, including consciousness, can arise from the collective activity of millions of neural components, each of which may be viewed as performing computational operations.

Hence, P&V's treatment misappropriates two central concepts in cognitive science to narrow and idiosyncratic definitions. It might be argued, though, that their position could be rescued by flagging this narrow usage with appropriate qualifiers. Thus, their central assertions could be recast as: (a) All semantic representations are conscious, and (b) there are no unconscious cognitive processes that operate upon semantic representations. This is a less confusing statement of their of their position, and it enables attention to be turned from matters of definition and conceptual analysis, to the more tangible arena of empirical evidence. Furthermore, both of these issues have been subjected to intense empirical scrutiny. The voluminous experimental psychology literature on semantic processing and consciousness is dealt with briefly in section 8.2 of the target article. Although space limitations preclude detailed discussion of this extensive body of work, two brief comments can be made. P&V quote Holender's (1986) negative conclusion on the status of unconscious semantic processing, but fail to mention that the vast majority of expert commentators on Holender's target article disagreed with this conclusion. Is this challenging the *Zeitgeist*, or ignoring the weight of evidence? On the other hand, P&V are correct in noting that since USP effects are typically weak and short-lived, this cautions against extravagant theorising concerning the functional role of unconscious semantic processes.

However, the most cogent grounds for accepting that complex cognitive processes can be dissociated from consciousness comes not from experimental psychology, but from neuropsychology. This literature contains a vast body of empirical and theoretical work discussing dissociations related to consciousness that have been observed in a wide variety of syndromes, including the split-brain syndrome, amnesia, deep dyslexia, prosopagnosia, and neglect. P&V devote just four paragraphs (sect. 8.3) to this body of knowledge, concentrating mainly on blindsight, and concluding with an interpretation of blindsight in terms of degraded performance accompanied by distorted phenomenal awareness. Weiskrantz (1997, pp. 36–48) examines this interpretation of blindsight in considerable detail, and provides a several cogent reasons for rejecting it (e.g., the presence of qualitatively different patterns of performance in blindsight and degraded normal performance). To be taken seriously as a framework for understanding consciousness, P&V's proposal will need to engage more fully important evidence and theory from neuropsychology concerning consciousness and cognitive processing.

In conclusion, although the framework proposed by P&V is admirably parsimonious, their analysis has two serious problems: (1) Their definitions of computation and of representation are flawed, and (2) their analysis and discussion fails to engage adequately with the neuropsychological literature concerning consciousness and cognition.

## Rules, abstractions, and evolution

Leonid Litman and Arthur S. Reber

*Department of Psychology, Brooklyn College of CUNY, Brooklyn, NY 11210.*
**psychdrum@yahoo.com areber@brooklyn.cuny.edu**

**Abstract:** Perruchet & Vinter's article, for all its breadth and scope, has several deep problems: specifically, an eccentric notion of rule, a narrow notion of what it means for a mental instantiation to be abstract, and a failure to take into account fundamental principles of evolutionary biology. While not the only problems, these three are sufficient to seriously weaken their arguments.

Perruchet & Vinter (P&V) present an intriguing and provocative view of human cognition built around the notion of a self-organizing consciousness. The role of the cognitive unconscious (an entity which, from their perspective, assumes nearly oxymoronic status) is relegated to the status of a rather stupid device that operates by rigid connectionist principles and is incapable of capturing rule-governed behavior or establishing abstract representations. There are many points of disagreement between this theoretical stance and ours, and more than a few issues need addressing. However, for the purpose of this commentary we will focus on three key issues: rules, abstraction, and evolution.

***1. On rules.*** The question of whether the unconscious follows rules that are instantiated symbolically or whether it functions on associative principles is, despite the intense focus P&V put on it, actually irrelevant. The fundamental questions really under the microscope here, are the complexity of the problems the unconscious can solve and the nature of the tasks the unconscious can perform. The particular strategy that evolved to solve those complex problems is a secondary problem.

While P&V maintain that the unconscious is neither powerful nor sophisticated, and that there is no such thing as an unconscious mental life, what they really seem to be inveighing against is not the power of the unconscious, but the way in which it carries out operations. Their argument revolves around the claim that rules are not coded symbolically in the unconscious, but rather are the products of a neural network that operates on principles of learning and associations. Alas, this move won't work; the power of a system and the way in which it carries out its computations are two separate issues.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

In what sense is it important that the unconscious is following rules rather than following a cascade of associations out of which the rules emerge? The confusion is most evident in their extended treatment of transfer in artificial grammar-learning tasks where they maintain that such transfer does not imply rule abstraction because it could be simulated by a connectionist network. P&V seem to think that claiming that the unconscious does not represent rules symbolically, and claiming that there is no such thing as a sophisticated unconscious mental life, are the same thing. But they clearly aren't. As one of us argued at considerable length (Reber 1993, pp. 111–20), rules are manifestations of the regularities of systems, be they connectionist in nature or captured by other architectures, and questions of how powerful, adaptive, and smart any given system is has more to do with the particular meta-psychological stance of the cognitive theorists peering down at the behavior of their subjects than on any in principle notions about the content of epistemic representations.

We also suspect that P&V's stance may be unsound on the grounds that it is not falsifiable. No matter how complex and powerful the unconscious is ever shown to be, they can always argue that the processes were "merely" connectionist in nature and therefore not rule-governed. Unhappily, P&V never tackle the deep issue: When is the behavior of a system sufficiently complex to warrant the conclusion that it must be following rules?

***2. On abstraction.*** P&V spent substantial intellectual coin on the notion that the representations of the unconscious are inflexible and nonabstract, arguing that the abstract and symbolic will only be found within the spotlight on the self-organizing consciousness.

Frankly, we think this notion is nothing but a red herring that detracts from the problem at hand. *Abstraction* denotes a continuum, not a feature of particular mental contents. The term itself comes from the Latin, meaning "drawn away from" (Reber & Reber 2001), and in some sense, all stimulus representations are abstract in the sense that their ultimate mental profile is distinct from the stimulus input. In Pavlovian conditioning, an organism originally trained with a CS of 1,000 Hz will show a CR to a 1,010 Hz tone. Priming with a word like *soldier* decreases RTs to words like *war.* Breaking functional fixedness while solving one problem facilitates working with novel problems requiring similar strategy. These are all examples of abstraction – the first being considerably less so than the latter two.

Why this issue seems so hard to grasp when topics like consciousness and abstraction are under discussion, escapes us. Psychologists have always treated memory in exactly this fashion. We are quite comfortable with detailed discussions of the memorial capacities of organisms from creatures as primitive as *Aplysia californica* to the more familiar and representationally sophisticated *Homo sapiens* without feeling any need to break the continuum into distinct categories such that one form of memorial content is "true" memory and another not.

Our point here is simply that it makes little sense to try to maintain that the explicit representations are abstract and the implicit are not. Indeed, as has been shown repeatedly (Manza & Reber 1997; Reber & Allen 1978; Whittlesea & Dorken 1993; Whittlesea & Wright 1997), identical implicit learning materials can yield instantiated or abstract representations depending on demand characteristics.

***3. On evolution.*** Interestingly, this topic received essentially no attention from P&V. Yet, it is of overarching importance. Any viable model of mind must, at the very least, be plausible within the framework of contemporary evolutionary biology. Given the perspective we have taken in the first two parts of this commentary, P&V's overall perspective fails this test.

Their point of view assumes that all observable rule-governed behavior and all abstract memorial representations are the products of a recently evolved consciousness (see Reber 1992a; 1992b; 1993; Reber & Allen 2000) and that no proto-rule functions or proto-abstract representations would be expected in other species or in our ancestors. Contemporary models of evolutionary biology simply do not allow for these kinds of rapid evolutionary saltations, certainly not within the time frame dictated.

It seems more reasonable to assume that systems that operated independently of the top-down, modulating, self-reflective (even self-organizing) consciousness that P&V are so focused upon, (1) emerged relatively early in evolutionary history, (2) functioned to provide veridical information so that organisms could operate effectively in a changing and highly variable ecological niche, and (3) enabled such organisms to capture critical features of their world in ways that allowed for appropriate flexibility of function (i.e., proto-rule-governed behavior) and generalization of action (i.e., emergent abstract representations).

Could such primitive systems have been based on connectionist nets similar to PARSER? Sure. Would such systems "have" rules and abstract representations? Sure, only they wouldn't be very complex or very abstract.

ACKNOWLEDGMENT

This commentary was prepared under support from NSF Grant No. 0113025.

## What does "isomorphism between conscious representations and the structure of the world" mean?

Riccardo Manzotti and Giulio Sandini

*LIRA-Lab, DIST, University of Genoa, Genova 16145, Italy.*
**manzotti@dist.unige.it** **sandini@dist.unige.it**
**http://manzotti.lira.dist.unige.it** **http://www.lira.dist.unige.it**

**Abstract:** Perruchet & Vinter's provocative article challenges a series of interesting issues, yet the concept of isomorphism is troublesome for a series of reasons: (1) isomorphism entails some sort of dualism; (2) isomorphism does not entail that a piece of the world is a representation; and (3) it is extremely difficult to provide an explanation about the nature of the relation of isomorphism.

Perruchet & Vinter's (P&V's) provocative article challenges a series of interesting issues: (1) that the "powerful cognitive unconscious" is an unnecessary hypothesis; (2) that most of the cognitive mental functions can be explained using elementary associative processes; and (3) that there is an isomorphism between the representations composing the phenomenal experience and the structure of the world. We could agree (with some differences) on the two apparently strongest claims, namely (1) and (2). However we challenge the very concept of "isomorphism" and the possibility of having data about what is considered to be "conscious activity."

The concept of isomorphism is troublesome for a series of reasons: (1) isomorphism entails some sort of dualism; (2) isomorphism does not entail that a piece of the world is a representation; and (3) it is extremely difficult to provide an explanation about the nature of the relation of isomorphism.

Let's start from the problem of dualism. By using the phrase *conscious representation,* P&V seem to imply that there is something different between being a representation and being a fact of the world, or between a "conscious representation and the unconscious processes generating those representations" (sect. 1.3.1). Hence the conscious mind should be something different, and as a result, P&V would need a proper ontology, an issue that they leave unexpectedly untouched.

As for the problem of the nature of a conscious representation, if a mental state is referred to as *conscious,* it must have some property that qualifies it as such. However, if a representation is defined as conscious only because it occurs as part of the mental activity of a conscious agent, it should be explained why the agent is conscious. P&V don't seem to provide any explanations of this. In the target article they refer to the difference between conscious and unconscious as a "status of the processed information" (sect.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

1.1). They refer to Dulany's work, which establishes the concept of representation on two cases: deliberative episodes and associations. However, they don't provide any rationale to maintain that deliberative episodes are sufficient to have true intentionality in the sense of aboutness (see Brentano 1874/1973; Dretske 1981; 1995; Dulany 1997; Harnad 1990; Searle 1983). P&V don't offer any direct explanation.

Finally, let's get to the problem offered by the concept of isomorphism. What does it means that "a representation has to be isomorphic to the world structure" (sect. 2.1.2), that "the phenomenal experience is isomorphic to the world" (sect. 2.2), and that "internal representations . . . are increasingly congruent with the world structure" (sect. 7.1.1)? Is it something similar to Ludwig Wittgenstein's "logical form"? Is it the syntactical similarity between what happens in the outside world and the internal relation between events? Is it something that can be obtained at different degrees and hence be quantified, as it would seem by the fact that "the initial conscious percept . . . becomes increasingly isomorphic with the structurally relevant units" (sect. 6.2)? Is it some kind of functional isomorphism? Other authors have tried to make use of the notion of isomorphism (O'Brien & Opie 1997), which is somewhat Cartesian because it entails that there's a mental copy of some property of the external world. For example, according to O'Brien & Opie, "we are required to think of representational content as a special kind of correspondence between intrinsic properties of neural activation pattern and aspects of the world," and "representation exploits a structural isomorphism between its physical substrate and its physical domain" (O'Brien & Opie 1999). However, up to now, nobody has presented any convincing example of what ought to be a convincing isomorphic representation of sensory modalities different from spatial relations (O'Brien & Opie 1999) and syntactical symbols (Perruchet & Vinter, target article). In fact, if a spatial relation can be somehow duplicated by some kind of spatial representation, how would it be possible to duplicate taste, smell, or other kinds of qualitative conscious contents (Manzotti & Sandini 2001; Manzotti & Tagliasco 2001)?

PARSER, the software model presented by P&V, is not convincing in this respect because it makes use of words as the input and of words as the output. In other words, PARSER seems to work in two congruent domains. In the real case, the two domains are not congruent at all. How can PARSER explain the semantic relation with colors, pains, emotional values, smells, flavors, and so on? Furthermore, in a sense, the inputs of PARSER are words only because we interpret them as such. In the same way, the information processed by a computational unconscious system are numbers only because they are interpreted as such.

Finally, we have some concerns about the availability of data on the nature of what P&V confidently call "conscious activity." How do we know that one process is better at mimicking conscious activity than another one? How do we know that PARSER is nearer to consciousness than other models? The obvious way is to refer to our personal first-person experience. Yet this move is highly dangerous because it trespasses the boundary between objective data and first person subjective reports. P&V write about "match[ing] the phenomenal experience" (sect. 1.2), "form[ing] conscious representations that are isomorphic to the world structure" (sect. 2), and "mimicking the on-line conscious processing of incoming information" (sect. 3.3). How can we get a feasible description of the nature and structure of our conscious activity, a description that is different from introspectionism? Present approaches don't offer any ready-to-use methods, not even Francisco Varela's second-person neurophenomenology, nor Daniel Dennett's heterophenomenology (cf. Dennett 1991; Thompson 2001; Varela 2000). This problem is crucial when it's time to evaluate PARSER results. If the authors themselves admit that "the functioning of PARSER . . . does not depend in any way on the conscious/unconscious status we ascribe to its components" (sect. 3.3), how can they claim that PARSER can help as a "model to simulate conscious states while respecting the properties of conscious thought" (sect. 3.3)? Does PARSER satisfy these properties? Are they among the relevant "properties of conscious thought" (sect. 3.3)? P&V don't say anything about it.

## The SOC framework and short-term memory

David J. Murray

*Department of Psychology, Queen's University, Kingston, Ontario, K7L 3N6 Canada.* **murrayd@psyc.queensu.ca**

**Abstract:** Using a particular formula for quantifying the effortlessness that Perruchet & Vinter suggest accompanies the detection of repetition among a set of representations concurrently in consciousness, it is shown that both the Sternberg function and the Cavanagh function, associated with immediate probed recognition tasks and memory span tasks, can be predicted.

Perruchet & Vinter (P&V) noted that the detection of a repetition within a set of representations concurrently in consciousness occurs "with little cognitive effort" (sect. 7.2), that is, rather effortlessly. In the recognition task studied by Sternberg (1966), a set of $L$ differing stimuli is presented successively, followed by an $(L + 1)$th stimulus that either is or is not a repetition of one of the $L$ stimuli. Sternberg found that the mean time to detect that the $(L + 1)$th stimulus is a repetition increases linearly with $L$, up to a limit of about $L = 7$. Denote by $a$ the slope of this linear "Sternberg function." Cavanagh (1972) discovered that $a$ itself is a linear function of the reciprocal of the memory span for the material being studied.

Let $m$ denote the probability that, in a Sternberg task, two differing adjacent list stimuli are each encoded into an identical, but partial, representation in consciousness. For example, the sequence of digit-triples "549 581" might be encoded by their initial digits only, as "5 5." According to the theory of Murray et al. (1998; 1999), whenever two adjacent mental representations are identical, the second overwrites the first.

In many Sternberg tasks, each stimulus is a *single* digit that is selected by random sampling *without* replacement from an ensemble. The lists used in memory span tasks also avoid repeated adjacent list stimuli (Crannell & Parrish 1957, p. 320; Woodworth & Schlosberg 1954, p. 696). Hence, in most Sternberg and memory span tasks, the value of $m$ will be low and that of $(1 - m)$ – the probability that two differing adjacent stimuli will be fully encoded as differing adjacent representations – high. When $(1 - m)$ is high, the total number of representations of list stimuli in consciousness, just prior to the $(L + 1)$th stimulus, approximates $L$ (Murray 2001).

We now consider the case where $L = 1$. The theory states that, if they are identical, the representation of the repetition should overwrite the representation of the single list stimulus. If not, the two representations probably differ on some dimension other than content. This dimension is probably temporal context, because Diener (1988) has shown that rising Sternberg functions are only obtained if the $(L + 1)$th stimulus is well separated from the final list stimulus.

We make two assumptions. First, if two adjacent stimuli identical in content differ in context, then the lower the probability will be that their corresponding representations are *exactly* identical. Second, the more these stimuli differ in context, the longer it will take to detect that their content has been repeated. Because a Sternberg function specifies that, when $L = 1$, the associated detection time is $a$, these two assumptions together predict that, as $m$ rises, $(1 - m)$ must fall, and $a$ should increase in duration:

$$da/dm = 1/(1 - m) \quad (1)$$

The solution of Equation 1 is

$$a = -\ln(1 - m) + u \quad (2)$$

where $u$ is a constant of integration.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

If the representation of the $(L + 1)$th stimulus is a repetition of a representation already in consciousness, there is no reason (apart from temporal context) why the representation of the repetition should not overwrite the representation already in consciousness. It was noted above that, before the $(L + 1)$th stimulus is added to a list of $L$ stimuli differing in content, the total number of representations in consciousness that differ in content approximates $L$. After the addition of an $(L + 1)$th stimulus that is not a repetition of a list stimulus' content, the total number of representations differing in content will be increased to $(L + 1)$. But if the $(L + 1)$th stimulus does repeat a content, and then the representation of that $(L + 1)$th stimulus overwrites the representation whose content it repeats, the total number of representations differing in content will remain at $L$.

Now, a participant would be helped to detect the fact that the $(L + 1)$th stimulus repeats a content if he or she notices that the addition of the $(L + 1)$th stimulus causes no observable change in the total number of *differing* contents represented in consciousness. Let P&V's effortlessness be denoted by $E$. We hypothesize that $E$ varies directly with the contrast between a change and no change in that number. So $E$ will be defined, relative to $L$, as:

$$E = [(L + 1)/L] - [(L + 0)/L] = (1/L) \quad (3)$$

Let $D$ denote the time required to detect that the $(L + 1)$th stimulus is a repetition, and let $D$ vary inversely with $E$. The greater the effortlessness, the shorter the detection time. Then:

$$D = (a/E) = [a/(1/L)] = La \quad (4)$$

The reason the constant of proportionality in Equation 4 is assumed to equal $a$ is that Equations 4 and 2 are thereby consistent because, when $L = 1$, $D = a$ in both equations. Furthermore, when $L > 1$, Equation 4 specifies that $D = (La)$. The term $(La)$ is conventionally called the "decision latency" component of the Sternberg function and is contrasted with a "residual latency" determined by routine response processes.

Next, we turn to the memory span task. Following the presentation of $L$ differing list stimuli, the participant must recall all $L$ stimuli in correct serial order. In a Sternberg task using digit-triple material, the probability that a representation of a particular list stimulus would still be in consciousness after the presentation of the final list stimulus, but prior to the onset of the $(L + 1)$th stimulus, was specified by Murray et al. (1998, p. 1199) to be $(1 - m)^i$. Here, $i$ is the number of list stimuli intervening between that list stimulus and the end of the list; usually $i = (L - 1)$. In a memory span task, the equivalent would be the number of stimuli (including both list stimuli and the recalls of those list stimuli) intervening between the presentation of a particular list stimulus and the recall of that stimulus. Here too, this number equals $(L - 1)$.

The probability, $P$, that *all* of the $L$ list stimuli will be correctly recalled in order, is:

$$P = (1 - m)^{L(L-1)} \quad (5)$$

Cavanagh (1972, p. 527) defined the memory span, $S$, as being that $L$ value associated with a probability of 0.5 of being correctly recalled in order. Dividing both sides of Equation 5 by $2P$, and substituting $S$ for $L$ in Equation 5, yields:

$$0.5 = [(1 - m)^{S(S-1)}]/(2P) \quad (6)$$

Taking natural logarithms and rearranging terms of Equation 6 gives:

$$-\ln(1 - m) = \{-\ln[0.5(2P)]/(S - 1)\}(1/S) \quad (7)$$

From Equation 2, the left side of Equation 7 equals $(a - u)$, so Equation 7 becomes:

$$a - [(-\ln P)/(S - 1)](1/S) + u \quad (8)$$

Equation 8 is Cavanagh's function expressing $a$ (the slope of the Sternberg function) as a linear function of $(1/S)$ (the reciprocal of the memory span). Its intercept is $u$ and its slope is $[(-\ln P)/(S - 1)]$. Cavanagh's estimated value of the intercept was 0.0028s, and his estimated value of the slope was 0.2432s.

Cavanagh also reported the obtained values of $a$ and of $S$ associated with each of seven materials. The estimated values of $m$ for the seven materials ranged from 0.0301 (for single digits) to 0.0678 (for nonsense syllables). The mean value of the seven estimates of the slope of the Cavanagh function, derived by calculating $[(-\ln P)/(S - 1)]$ for each material, was 0.2431s.

ACKNOWLEDGMENTS

This research was funded by Social Sciences and Humanities Research Council of Canada Research Grant 410-99-0831. I am most grateful to Rachel Murray for her help in preparing this manuscript as a plain text HTML file, and to Christina Bandomir, Geert-Jan Boudewijnse, Sylvie Hebért, Andrea Kilgour, Kyunghwa Kwak, and Louise Wasylkiw for their support during the exploratory phases of this investigation.

## The computational baby, the classical bathwater, and the middle way

Gerard O'Brien and Jon Opie

*Department of Philosophy, University of Adelaide, South Australia 5005.*
**gerard.obrien@adelaide.edu.au** **jon.opie@adelaide.edu.au**
**http://www.arts.adelaide.edu.au/philosophy/gobrien.htm**
**http://www.arts.adelaide.edu.au/philosophy/jopie.htm**

**Abstract:** We are sympathetic with the broad aims of Perruchet & Vinter's "mentalistic" framework. But it is implausible to claim, as they do, that human cognition can be understood without recourse to unconsciously represented information. In our view, this strategy forsakes the only available mechanistic understanding of intelligent behaviour. Our purpose here is to plot a course midway between the classical unconscious and Perruchet &Vinter's own noncomputational associationism.

We are sympathetic with the general thrust of Perruchet & Vinter's (P&V's) "mentalistic" framework. In particular, we endorse their attempt to establish a principled connection between mental representation and conscious experience (see, e.g., O'Brien & Opie 1999a). And, like them, we are suspicious of orthodox, intellectualist approaches to the cognitive unconscious. Nonetheless, in developing the notion of a "self-organizing consciousness," P&V go too far when they contend that unconsciously represented information plays no role in human cognition. As we explain shortly, this thesis throws out the computational baby with the classical bathwater, and thereby forsakes the only available mechanistic understanding of intelligent behaviour. Our purpose in this commentary is to plot a course midway between the classical unconscious and P&V's own noncomputational associationism.

At the heart of P&V's project is the claim that (unconscious, noncomputational) associative processes of learning and memory are sufficient to generate conscious mental representations whose contents are isomorphic with the world. This is a very bold claim. It flies in the face of the classical computational program in cognitive science, which assumes that conscious representations are the product of a vast amount of unconscious symbol manipulation governed by unconscious rules. It is therefore incumbent on P&V to provide a detailed account of the kinds of associative processes they have in mind.

Yet, as far as we can determine, P&V directly address the nature of mental association in just three relatively short passages of the target article. First, they briefly discuss how associative learning can explain classical conditioning (sect.1.3.3). Second, in response to the apparent inability of associative processes to account for the full complexity of human cognition and conscious mental life, they suggest that such processes are not limited to acting on elementary stimuli but can be extended to complex representations (sect. 2.2.2). And third, they claim that primitive represen-


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

tations that are repeatedly linked in attention "tend to pool together and form a new primitive," which is then available as a unitary representation in subsequent processing (sect. 3.2). All of the ensuing discussion presupposes that mental association has been adequately explained, and P&V concentrate on showing how the complex conscious representations it produces can account for human behaviour in various domains.

What we are suggesting is that P&V's discussion of mental association is insufficiently developed to bear the burden of their principal thesis. This becomes especially apparent in section 6.4, where P&V help themselves to representations with increasingly abstract contents. These, they explain, can be formed by iteration of the same basic associative processes already described. Yet, because they have said so little about these processes, it is difficult for the reader to assess the plausibility of this claim.

It is not surprising that P&V fail to specify associative processes adequate to the complexities of human cognition. Associationism, in various guises, has been around for a long time, and its shortcomings are well known. Two of these deserve mention here.

The first concerns the raw materials on which associative processes putatively operate, specifically, conscious representations. Whatever the status of the claim that the conscious states produced in the course of processing linguistic input are the result of associative processes,[1] the same cannot be said of other perceptual modalities, especially vision. This is because visual experience is radically underdetermined by the data available in the proximal stimulus (the dynamic array of light falling on the retinae). The conclusion most theorists find inescapable is that visual perception is an inferential, rather than associative, process – one that constructs visual representations by combining stimulus data with internally represented assumptions about the structure of the environment. Because we are not aware of this process, visual inference, and the representations it implicates, must be unconscious (see Fodor & Pylyshyn 1981; and for a more recent statement, Palmer 1999, pp. 55–59).

The second problem runs even deeper and concerns the physical implementation of cognitive processes. What is required to explain cognition is a physical mechanism that can account for the parallelism between the content of mental representations and their causal relations (see, e.g., Fodor 1987, Ch.1). This challenge stumped proponents of associationism in the eighteenth and nineteenth centuries, who only permitted contiguity and co-occurrence of ideas as ultimate explanatory principles. Behaviourism, essentially a nonmentalistic version of associationism, was likewise explanatorily impoverished. It was only the emergence of the classical computational theory of mind, inspired by the power of digital computers, that saw the first serious contender for a mechanistic account of cognition. Classicism offers a story as to how the causal interactions between mental representations (in the form of symbol structures) preserve their semantic relations. But this story is distinctly nonassociationist, given that it depends on the operation of unconscious, structure-sensitive rules. Because these rules, and much of the information they implicate, are not consciously available, classicism delivers the sophisticated cognitive unconscious P&V are at pains to avoid.

In short, as things currently stand, the only idea we have of how cognitive processes might be physically implemented in the brain assumes that these processes are computational in form. To abandon computationalism in favour of a mentalistic form of associationism, as P&V exhort, is thus to embrace mystery.

But P&V need not despair. There is a path that runs midway between classical computationalism and mentalistic associationism; a path that avoids the classical unconscious *without* abandoning computationalism. We finish by plotting this middle way.

The middle way is to reject classicism, with its commitment to a symbol crunching unconscious, in favour of the *connectionist* computational theory of mind. As we have explained at length in this journal (O'Brien & Opie 1999a), connectionism permits a distinction between explicit (activation pattern) representation and inexplicit (connection weight) representation that is perfectly crafted for P&V's mentalistic framework. Instead of claiming that unconscious information plays no role whatever in human cognition, they should restrict this claim to unconscious, *explicitly represented* information. Even if all explicitly represented information is conscious, inexplicit representation is still free to carry the burden of the cognitive unconscious. In this story, the links between conscious representations are computational, rather than merely associative, because they are mediated by connection weight representations. Such representations embody, for example, unconscious assumptions about the visual world, but in nonsymbolic form. Thus, the unconscious one ends up with is very different from its classical counterpart. Moreover, the conscious (explicit) and unconscious (inexplicit) are intrinsically and deeply connected, because activation pattern representations are shaped by inter-unit connections. To use P&V's own metaphor (sect. 1.3.1), they are like the two sides of a coin, inseparable, yet distinct.

NOTE

**1.** P&V's work on this topic is, we think, highly significant, and does much to undermine traditional models of word segmentation, grammar learning, and so on, which help themselves to an implausibly rich cognitive unconscious.

## Oral and visual language are not processed in like fashion: Constraints on the products of the SOC

Christophe Parisse[a] and Henri Cohen[b]

[a]*INSERM-Laboratoire de Neuropsychopathologie du Langage et de la Cognition, Batiment Pharmacie, Hopital de la Salpetriere, Paris, France;* [b]*Cognitive Neuroscience Center and Department of Psychology, Université du Québec à Montréal, Montreal, QC, Canada.*

**parisse@ext.jussieu.fr henri.cohen@uqam.ca**

**http://www.plpc.chups.jussieu.fr/christophe.htm**

**http://www.unites.uqam.ca/sirp/experts/6593.html**

**Abstract:** The SOC framework does not take into account the fact that the oral modality consists of purely transient data, which is not the case for the other modalities. This, however, has important consequences on the nature of oral and written language, on language consciousness, on child language development, and on the history of linguistics.

In section 2.1, Perruchet & Vinter (P&V) posit the existence of an isomorphism between "the actual and the represented world," and explain that "complex representations account for seemingly rule-governed behavior." This is made possible by the existence of the self-organizing consciousness (SOC), the principles of which are exemplified in PARSER. The power of PARSER was put to the test on a computer by replicating a learning situation tested with infants and adults by Saffran and collaborators (Saffran et al. 1996a; 1997). The situation corresponds to the extraction of words from raw phonetic input.

The performance of PARSER is very impressive – as is that of infants and adults – but maybe not so much so when one considers the characteristics of the input signal. All syllables in the input are of the same duration, the same height, share the same structure, and are repeated more than a thousand times in the same situation. In such contexts, primitives such as syllable segmentation and syllable identification work so well that the problem to be solved becomes too simple and does not adequately represent real life situations. Of course, there is a lot of information other than word regularities that helps children get the job done in real life situations (see target article, sect. 4.1; Johnson & Jusczyk 2001; Perruchet & Vinter 1998b). Nonetheless, this affects the working context of PARSER and undermines the demonstration of P&V.

A second problem is that stimuli in the oral modality are always transient. This limits the possibility of an "outside memory," (O'Regan 1992) and makes it more difficult to create an isomorphic representation. Conscious reanalysis of the signal is impossible or



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

very constrained by our linguistic knowledge (we replay the signal as we decoded it; we cannot look at it again), unless the words or the strings of words are repeated numerous times in the same context, which is unlikely in real-life situations. One way to circumvent this limitation is through the sensori-motor loop of language production, but children do not repeat or produce all the words that they understand, so this is obviously not the only route into language structure.

However, we think that these limitations do not invalidate PARSER and the principle of SOC, if one takes into account the differences between the oral modality and other modalities. For other modalities, data are often not transient and it is possible to take full advantage of SOC to use the world as an "outside memory" and to trade representation against computation. Visual presentations and tactile sensations are much more likely than oral stimuli to lead to complex representations. Now, what is true for oral language is not true for written language. Thanks to the visual format of written language, it becomes much easier to build conscious representations isomorphic to the outside world. What necessitated thousands of syllables for oral language (Perruchet & Vinter 1998b) can be accomplished with much fewer repetitions. This makes it easier to analyze language per se, further disconnected from its semantic interpretation, than it was for oral language.

Our proposal is that SOC works with both oral and written language but results in different products, and that structural differences between these products have great impact on the development of language in the child and on the nature and structure of language.

The characteristics of oral language make it difficult for young children to analyze the oral signal into parts, although they are already able to understand and manipulate a great number of words. Children actually take quite a long time before being able to split into parts and manipulate the oral forms that they used as unanalyzed wholes (Peters 1983) and exhibit knowledge of syntactic categories (Tomasello 2000a). However, the greater permanence of visual and tactile representations allows children to use these representations as a support for their first oral stimuli representations. This is consonant with their item-based linguistic behavior (Tomasello 2000b) and with the importance of the semantic nature of their first linguistic knowledge. Following the preschool years, the structure of written language becomes the outside support for internal isomorphic representations, which become in turn a support for reasoning about language. Indeed, metalinguistic thinking appears when children have learned to read, not before (Gombert 1992).

As representations can be much more complex with written material than with oral material, real or seemingly rule-governed behavior would be much more developed in written language than in oral language. This can explain a lot about the differences between oral and written language. These differences can be found in the grammatical structures produced by native speakers (Miller & Weinert 1998), or in the structures and interpretations of texts and stories (Ong 1982). Oral language tends to be more formulaic than written language and mostly composed of first-order structures. Also, oral language structures and representations are more semantics-centered than syntax-centered with a lot of sometimes very complex, item-based constructions (see Croft 2001), as is the case for preschool children (Tomasello 2000b).

Finally, in the history of the human civilizations, the metaanalysis of language always appeared after the discovery of writing (Auroux 1994). The first written productions always were ideographic and only later became phonologic through a lengthy historical process, which eventually led to the development of the linguistic sciences (see Auroux 1994; Harris 1980). It seems as if a visual support was a necessary prerequisite in the detailed elaboration of the structure of the oral support and help in the transition from a purely semantic-centered view of language to a purely syntactic-centered view.

Our conclusion follows the proposal of Auroux (1994), that there are different levels of language consciousness: first, epilinguistic knowledge (the unconscious knowledge of language), which is a pure product of oral language; and, second, metalinguistic knowledge (the conscious knowledge of language), which is a byproduct of the use of written language. This opposition is unfortunately not yet taken into account in most linguistic and psycholinguistic theories, but it could adequately be described and explained within the SOC framework.

## Neo-associativism: Limited learning transfer without binding symbol representations

Steven Phillips

*Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan.* **steve@ni.aist.go.jp**
**http://www.etl.go.jp/~stevep**

**Abstract:** Perruchet & Vinter claim that with the additional capacity to determine whether two arbitrary stimuli are the same or different, their association-based PARSER model is sufficient to account for learning transfer. This claim overstates the generalization capacity of perceptual versus nonperceptual (symbolic) relational processes. An example shows why some types of learning transfer also require the capacity to bind arbitrary representations to nonperceptual relational symbols.

Perruchet & Vinter (P&V) claim that with the addition of a simple principle, the capacity to determine whether any two arbitrary stimuli are the same or different, their association-based PARSER architecture is sufficient to explain learning transfer, such as that reported by Marcus et al. (1999). Learning transfer relies on between-stimulus relations that are common across different sets of unique stimuli. However, the concept of a common relation belies an important distinction between perceptual and nonperceptual (symbolic) relations that has different implications for generalization capacity. Consequently, although their proposal may be sufficient to explain, for example, transfer in infants based on common perceptual relations; it is not sufficient for transfer based on common nonperceptual relations. And, by extension, that infants demonstrate transfer on a phoneme sequence recognition task, does not imply the capacity to process symbolic relations.

***Perceptual versus nonperceptual (symbolic) relations.*** A perceptual relation is computed from the perceived features of the related stimuli. *Sameness/difference* is a perceptual relation where there exists a thresholded distance function over any two points in feature space. Computing perceptual relations affords learning transfer in relation-match-to-sample, because the task is reduced to learning that a match is found when the distances between paired stimuli in source and target pairs are either both less than, or both greater than threshold. Because matching perceptual sameness/difference relations is the essence of the tasks analyzed, I agree with the authors that PARSER is capable of transfer on these tasks. By contrast, a nonperceptual (or, symbolic) relation, such as *sibling*, is one where no such function over perceptual features is available.

Of itself, this distinction is unremarkable. But, in the context of characterizing cognitive processes it lies at the root of the problem which appears in the authors' claims and the sorts of data they are attempting to explain. Perceptual relations can be *imposed* on nonperceptual relations by prior learning of new perceptual relations, and these learned perceptual relations can be used as the basis for transfer (e.g., Hinton 1990). The problem for the authors (and this type of explanation, generally) is how much of this prior learning justifiably explains learning transfer. It seems reasonable to suppose that infants already have the capacity to distinguish arbitrary pairs of phonemes prior to the phoneme sequence recognition task. But, for other tasks, such as Hinton's Family Tree and Transverse Patterning, no amount of prior learning justifies learning transfer in the networks analyzed (Phillips 1999; 2000). If computing sameness/difference perceptual relations is sufficient, how



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

does one explain transfer when this sort of information is not available? A nonperceptual (symbol) version of relation-match-to-sample is defined to illustrate the problem and the limits of models sensitive only to perceptual relations, including PARSER.

***Symbol-relation-match-to-sample.*** In this task, the subject is presented with four pairs of stimuli: AB, CD, EF, and GH. Symbol-relation-match-to-sample differs from relation-match-to-sample in that the sameness/difference relation is specified by colored backgrounds, not features specific to the stimuli. Here, brackets indicate colors, so that {A}{B} – identical background colors – identifies the symbolic relation SAME (A, B); and {A} [B] – different background colors – identifies the symbolic relation DIFFERENT (A, B). The subject is then presented with an uncolored probe pair and the three remaining uncolored choice pairs. Subjects must select the choice pair with the same symbolic relation as the probe pair. For example, given pairs {A}{B}, {C}{D}, {E}[F], and {G}[H]; uncolored probe pair AB; and uncolored choice pairs CD, EF, and GH: The correct response is CD. For probe EF, the correct response is GH.

The critical aspect of this task is that the correct response cannot be based on any perceptual relation between the uncolored paired stimuli, because all uncolored stimuli are different, and any coincidental perceptual relation used to predict the response in one trial would fail on a subsequent trial where the pair were reassigned to a different symbolic relation. Thus, no amount of prior learning on perceptual relations can account for transfer on this task. Transfer is achieved with the capacity to bind arbitrary pairs of stimuli to symbols (representing the symbolic relations SAME and DIFFERENT); and depending on the decision process, either the capacity to match retrieved pairs from relational symbols or match retrieved relational symbols from pairs.

Marcus (1998) has argued that cognitive processes must be symbolic, because only symbol systems can explain the capacity to generalize across novel stimuli. Yet, the infant data used as evidence of symbol systems are open to two different interpretations. This comment is not to deny the importance of the authors' model or Marcus et al.'s infant data. To the contrary, they have helped us formulate a tripartite theory of cognitive processes, where this type of generalization behavior lies at an intermediate level between symbolically structured and nonstructured elemental associative processes (Halford et al., submitted).

## Associative learning: A generalisation too far

Martin Redington

*Department of Psychology, University College London, London, United Kingdom.* **m.redington@ucl.ac.uk**
**http://www.homepages.ucl.ac.uk/~ucjtfmr/**

**Abstract:** I argue that Perruchet & Vinter's claim that representations are conscious, and processes unconscious, gives too much ground to the cognitive unconscious; and that the boundary between conscious and unconscious mental phenomena is unlikely to fall neatly along these lines. I also propose that in the absence of more detailed models that demonstrably provide a reasonable account of the data, claims that associative mechanisms may underlie all cognition are premature.

I am greatly sympathetic to the general thrust of the argument advanced by Perruchet & Vinter (P&V) and in close agreement with their interpretation of much of the evidence that they cite.

For implicit learning, associative mechanisms provide the best available account of human performance, including the transfer phenomenon. There is no convincing evidence for rule learning, or for the existence of unconscious knowledge.

For language acquisition, structural information in the input, accessible to associative learning mechanisms, contains a great deal of information about linguistic structure, which is potentially available to the learner.

Given the success of associative learning mechanisms in these two domains, one may reasonably speculate, as P&V have done, that similar mechanisms may underlie many other aspects of cognition.

However, P&V's thesis goes far beyond this, making admirably bold claims about the accessibility of representations and processes to conscious awareness, and on the nature of learning mechanisms underlying human cognition generally. I believe that these generalisations are flawed.

First, consider the proposal that all representations are consciously accessible and the processes that generate them are unconscious (where representations are mental events that "assume the function of some meaningful component of the world" [target article, sect. 1.3.2] such as a person or object).

Although computational models have generally neglected issues of conscious accessibility, many models can be read as compatible with P&V's claim, with attention ensuring that the appropriate input is presented to a mechanism whose inner workings are inaccessible to consciousness, but whose outputs are available for further processing or action. This is especially true of connectionist models, where representations (patterns of activations over units) are transient, as in P&V's account, occurring only in response to appropriate inputs.

This compatibility suggests that P&V's claim may be a useful characterisation of some aspects of cognition, but as a general account, ironically, their position surrenders far too much ground to the cognitive unconscious. If some processing is conscious, as P&V admit, then, given their argument on representations (some representations are clearly conscious, so it is more parsimonious to assume that are no separate unconscious representations), why assume a separate class of nondeliberative and wholly unconscious processing? This assumption conflicts with Shanks and St. John's (1994) view, that in many domains, and especially implicit learning, knowledge acquisition is the result of conscious processes acting on conscious representations. While I share the skepticism of both sets of authors about many claims for the cognitive unconscious, there seem to be many mental events to which we may in principle have access but generally do not attend to, and others (e.g., neural events) to which we uncontroversially have no conscious access. I do not see a strong case for supposing that the boundary separating accessible and inaccessible mental phenomena falls neatly between representation and processing.

Second, P&V claim that associative learning and memory mechanisms are sufficient to account for human cognition, and that rule-based mechanisms are not required. I consider this separately, because one can imagine variants of P&V's account that argue that representations are conscious but permit them to be symbolic, or conversely, that rely only on associative processes, while admitting unconscious representations or knowledge.

P&V's claim is compatible with the evidence from implicit learning, but is much more problematic when applied to language acquisition. Undoubtedly, structural properties of language contain a lot of information about some aspects of linguistic structure, despite influential, but mistaken, a priori dismissals (e.g., Pinker 1984). For example, in Redington et al.'s (1998) analysis of child-directed language, word co-occurrence statistics were highly correlated with a word's syntactic category. The availability of information in the input is no guarantee that it is utilised by the learner, but in this case the cues are so informative, and so readily accessible to simple learning mechanisms, that it would be very surprising if learners did not exploit them in some way.

Nevertheless, distributional properties are far from perfectly informative about syntactic category, and it seems likely that additional cues, such as semantic and phonological constraints, prosodic information, and innate knowledge (for example, of the number and/or relative proportions of words of each category), must also contribute. While one may speculate, as P&V do, that associative mechanisms may also be able to exploit semantics, prosody, or phonology (leaving aside innate constraints), and to appropriately combine this information to identify syntactic cate-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

gories, in the absence of both a more detailed account of how information is extracted and integrated from these various sources and models that capture the facts of acquisition, claims that associative processes are sufficient are premature.

A similar objection applies to P&V's conclusions from their studies of the word extraction (segmentation) problem using their PARSER model. Although chunking strategies (such as PARSER, or Wolff's 1977 model, or Servan-Schreiber & Anderson's 1990 competitive chunking model) are able to extract some information about word boundaries from natural language input, Brent and Cartwright's (1996) minimum description length model provides superior performance on natural language inputs. Although P&V neglect to provide any comparisons, my own exploration of the chunking models suggests that PARSER's performance will be inferior to Brent and Cartwright's model. P&V argue that biological plausibility favors their model, but it is unclear how biological plausibilty of one model should be traded off against the superior performance of another. In the absence of a purely associative model that performs as well or better, the Brent and Cartwright model cannot simply be dismissed.

P&V's case weakens further when we consider more complex aspects of language, such as grammar. Although symbolic accounts of linguistic knowledge are sometimes vague, and symbolic computational models of acquisition or performance are disappointingly absent, symbolic accounts of grammar possess far more explanatory power than any associative challenger. Although much of this explanatory power may be attributed to the close relation between these accounts and symbolic descriptions of language derived from linguistics, as P&V suggest, they still represent a far better account of grammatical competence than P&V's alternative.

Advocates of symbolic rule-based accounts of language were mistaken in dismissing the potential contribution of distributional information to any aspect of language acquisition without examining the actual evidence. P&V are mistaken in generalising from the utility of distributional information and associative learning mechanisms in some aspects of language acquisition to the position that such mechanisms could account for all linguistic knowledge, in the absence of any compelling account of how associative mechanisms can account for the facts of linguistic performance.

P&V's challenge to assumptions of the cognitive unconscious and rule-acquisition should be applauded, and provides useful insights into possible relationships between learning mechanisms, representation and awareness, but only careful comparison of the evidence for competing accounts (and ideally comparison of computational models) of learning in specific domains will allow us to tease apart these issues. I suspect that the picture revealed by such comparisons will turn out to be far messier and more complex that the one portrayed by P&V.

## Is syntax a representation in itself?

Maritza Rivera-Gaxiola and Juan Felipe Silva-Pereyra

*Speech and Hearing Sciences Department, Center for Mind, Brain and Learning, University of Washington, Seattle, WA 98119.*

**rivegaxi@u.washington.edu** **jfsp@u.washington.edu**

**Abstract:** We address the issue of the nature of representations during development regarding language acquisition. The controversy of syntax as a process or operation for representation formation and syntax as a representation in itself is discussed. Eliminating the cognitive unconscious does not warrant a simplified, more parsimonious approach to human cognition in general.

Perruchet & Vinter (P&V) distinguish between processes or operations whose end product is the conscious representation of an immediate event, and the self-organizing consciousness or conscious representations themselves. According to their view, processes or operations need not be conscious, but they manipulate representations with a conscious content. The validity of these assumptions is evaluated, among others, for language processing.

We believe that the general argument of a self-organizing consciousness is feasible. However, there are some caveats that make the framework incomplete: First of all, although mentioned in the text, it is not clear or consistent throughout the article what is the true "raw" material with which an unconscious operation works. If unconscious operations are an intrinsic part of the two-types-of-events mental life claimed by the authors, the integration of processes and representations remains unclear. Moreover, it is unclear to what extent operations are unconscious because the mentalistic framework restricts formation and manipulation of abstract knowledge to conscious activities.

The authors assume that there is an isomorphism between our mental representations and the structure of the world. In their view, representations are defined as *mental events that assume the functions of some meaningful component of the represented world.* Taking a developmental perspective, one may ask what is *"meaningful"* to an infant? How do events become meaningful in the infant's mind? If consciousness self-organizes progressively, relying on the subject's immediate experience, it becomes difficult to accept the notion that the infant's isomorphism is the same as the adult's, or for that matter, the isomorphism of one adult with respect to any other. We therefore question their intuitive concept of isomorphism altogether.

In a more particular fashion, we address the question of whether syntax is a process or a representation in itself. In the authors' view, language acquisition rests on the basic principles of associative learning. Thus, the human being in his early life learns the grammar of his own language, or language to-be, using the repetitions and regularities of linguistic events to which he is exposed. He creates generalizations about the relationships between the representations being formed and understands what is being told and/or what he actually says. In our view, this can be a reasonable explanation were one to employ grammars with a small finite number of elements, such as artificial grammars. In natural grammar, however, the number of elements and combinations to which one is exposed does not allow resolving the understanding and producing of sentences and discourse never previously heard or represented. The age-old argument of the "poverty of stimulus" comes immediately to mind.

If syntax is a representation, or an abstraction in itself, and it is a product of associative learning, then we should push the assumption to its last consequences and wonder why other cognitively complex species perfectly capable of rule abstraction and associative learning do not reach the syntactic levels of performance expected, given their heavy training with symbolic systems that are also very rich in regularities. If human language intrinsically contains the syntactic representation, and the processing system in itself does not need knowledge of the abstract rules that govern the order of linguistic elements, what are we missing in the picture?

Nevertheless, if we consider syntax an unconscious operation for representation formation, we may be guided towards the traditional answer to the puzzle of language acquisition: There are, to say the least, prior perceptual and analytical constraints in the child's brain. The syntactic operation cannot be guided by the principles of associative manipulation of abstract knowledge, but exactly the opposite: The syntactic operator guides the way in which associations may take place. Selective prior constraints would guide the accommodation of elements in specific ways, such that the relationships of word and phrase order would become meaningful representations, regardless of having been exposed to them before and regardless of the content words having a meaning (see the following paragraph). The series of linguistic events, with all their regularities and irregularities, would be then unconsciously processed in selective ways. The world is structured, the mind construes it.

Recently, Hahne and Jescheniak (2001) showed that adult event-related potentials (ERPs) to sentences with and without



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

phrase structure violations are different regarding the amplitude of a specific component (P600). The authors showed the effect even in jabberwocky sentences (i.e., sentences in which content words have been replaced by pseudowords, while retaining morphological markers). In their case, subjects had to carry out a grammaticality judgment; however, preliminary results obtained in our lab show the same tendency in 3-year-old children who, in our case, are passively listening to the jabberwocky and non-jabberwocky sentences with and without syntactic violations presented (Silva-Pereyra et al. (2003). We therefore conclude that, even if semantic content is partially eliminated, the system will be able to detect syntactic violations. The momentary phenomenal experience consists of a syntactic operation without providing any final meaningful representation.

The model as presented by the authors is not simpler or more parsimonious than accepting the cognitive unconscious. It appears more a matter of gradation and terminology than a completely new framework. The issues of cognition in general and of language as a structure, as a mode of action as well as the origin of human language in particular, cannot be sufficiently explained by attributing human mental life to conscious activities alone.

## Does the SOC theory avoid unconscious rule use?

Carol A. Seger

*Department of Psychology, Colorado State University, Fort Collins, CO 80523.* **seger@lamar.colostate.edu**
**http://psyweb.psych.colostate.edu/psylist/detail.asp?Num=2**

**Abstract:** There are striking similarities between Perruchet & Vinter's SOC theory and the theories that they claim to be in opposition to. First, both kinds of theory lead to the induction of knowledge of complex patterns in the environment that serve the function of rules. Second, both involve unconscious mental processes of induction resulting in at least some conscious knowledge.

Perruchet & Vinter (P&V) propose their self-organizing consciousness (SOC) theory and related PARSER model as an existence proof that learning is possible without involving unconscious rules. They argue that belief in unconscious rule learning is widespread in cognitive science: For example, their target article begins: "The conventional cognitive framework rests on the existence of a powerful cognitive unconscious." Thus, it is fair to ask whether their theory truly avoids the use of rules within a powerful and sophisticated cognitive unconscious.

To begin, let's consider the definition of a rule. Within the cognitive sciences, the term *rule learning* is used to refer to at least two different kinds of mental representation formation. One use of the term follows the natural language meaning of rule, by referring to the induction of verbalizable rules formed via a process of conscious hypothesis testing. This form of rule learning was the focus of the classic early cognitive psychology research of Bruner (Bruner et al. 1956) and Levine (1975), and has seen a resurgence of interest in functional neuroimaging studies in recent years (Smith et al. 1998). In this research, subjects are typically taught to differentiate between concept exemplars on the basis of a verbalizable rule. For example, "blue items are in concept A" or "red and large items are in concept B." These studies demonstrate that subjects do generate and test verbalizable rules. This definition clearly does not lead to any learning that would be considered problematic by P&V: The verbalizable rules are completely available to conscious inspection and report.

Another use of the term *rule learning* is to refer to the induction of knowledge about regularities in the world, particularly in the field of language. Thus, people are said to have knowledge of the rules of the grammar of their native language, even if they are unable to verbalize these rules or otherwise have them consciously available. This meaning of rule is apparently what P&V intended to challenge, as is demonstrated by their emphasis on the domain of language.

What is the alternative that P&V offer to unconscious powerful rule induction mechanisms? In their SOC theory, features of the environment are mentally coded by perceptual mechanisms. Features that are attended to simultaneously are subjected to associative learning mechanisms that induce the existence of a relationship between the features and form a mental representation of the association. This process is hierarchical and recursive, so that lower-level associations between features can themselves become features that can be associated with each other. This system can abstract patterns existent in the world and apply them to new stimuli. For example, the letter string RMR contains a "trill" in that the first and last letter are the same. Once the pattern of a trill is learned, it can be transferred to a new letter set, such as VQV, as described in section 6.3. In sum, the SOC theory can find regularities in the environment, can represent them in an abstract manner, and can transfer the knowledge of these regularities to new situations that have different surface features. These are the functions of rules in the theories that P&V claim to be different from rules. However, if the mental representations they propose act like rules, then shouldn't P&V consider them to be rules? How would we ever show the difference experimentally?

Furthermore, the SOC theory proposes that the mental representations of features (both low level and high level) are accessible to consciousness, but the associative mechanisms that elaborate on these representations are "intrinsically unconscious." So, how does the SOC theory differ from the alternative theories that P&V claim assume a "powerful cognitive unconscious"? Both SOC and the alternatives include the presence of processes that are unavailable to consciousness and have a high degree of inductive power. The only apparent difference is that P&V assert that mental representations in the SOC theory are always consciously available, whereas some researchers allow some mental representations to be unconscious in the alternative theories. It is important to note that most researchers who postulate unconscious knowledge fully acknowledge that conscious knowledge is acquired simultaneously (Seger 1997) and do not propose that all knowledge is unconscious. However, the dividing line between an unconscious process working on a conscious representation is inherently slippery because of the possibility of structure-process tradeoffs (Anderson 1978), making P&V's assertion untestable. It would be all too easy to make their theory account for any data by taking a circular approach. Is it conscious? must be part of the representation. Is it unconscious? must be part of the processes working on the representation. Thus, it is unclear how different in practice P&V's theory is from the alternatives.

## Remember the old masters!

Benny Shanon

*Department of Psychology, The Hebrew University, Jerusalem, Israel.*
**msshanon@mscc.huji.ac.il**

**Abstract:** Perruchet & Vinter (P&V) ground their arguments in a view they call "the mentalistic tradition." Here I point out that such a view has already been advocated by two old masters of psychological science, William James and James Gibson, as well as by the philosopher Merleau-Ponty. In fact, in the writings of these older thinkers, arguments very similar to those presented in the target article are found.

I am very sympathetic to the view presented by Perruchet & Vinter (P&V) in the target article, which I found engaging and very well argued. Indeed, I too have argued against the founding of psychological explanation on hypothetical covert underlying mental representations and have proposed that, by and large, the domain of the psychological is coextensive with the domain of the



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

conscious (see Shanon 1993). In this commentary I would like, however, to mention what seem to me to be significant omissions in the target article. These omissions involve two great masters of the science of psychology, William James and James Gibson, who (had they been around) would have, I surmise, applauded this article, but also, I suspect, have said, "Sure enough, the arguments for this are already found in my own writing." (As shall be apparent, what I have in mind are specific argument patterns. By no means am I saying that everything presented in the target article has already been conceived by the earlier thinkers. Not at all!)

P&V proclaim themselves to be mentalists. By this, they mean that they adopt the view whereby the domain of mind is coextensive with that of the conscious, and that in order to model cognition, unconscious structures and processes are not needed. The authors associate this mentalistic line with the intellectual tradition of Descartes, the philosopher. In modern psychology, the mentalist par excellence is William James. A couple of years before the advent of psychoanalysis and the Freudian unconscious, and decades before the representational view of mind in cognitive psychology, James explicitly equated the mental with the conscious and launched a detailed critique of psychological explanations involving the postulation of unconscious underlying mental representations. In Chapter 6 of his *Principles of Psychology* (James 1890/1950), he examines ten arguments for the putative postulation of unconscious mental structures and processes and presents arguments against all of them. Two points feature centrally in James's argument. The first centers on a conceptual distinction (for James, a fundamental one) between "brain facts" and "mind facts." If psychological phenomena can be directly explained in terms pertaining to the brain, no additional covert psychological structures and processes have to be invoked to account for them. The second point has to do with the alternative explanation of empirical facts. James points out that in many cases where the phenomena are characterized as unconscious, there actually was a conscious event; only it was very short, faint, or not attended to. Consequently, the event was forgotten, and no credit to its existence is given.

Gibson (1966a; 1979) was for many years a solitary, ardent critic of representationalism. His theory of direct perception is grounded in the claim that for perception (as well as other cognitive functions) no mediating representational structures are needed. Along with several other observations allowing Gibson and his followers to dispense with representations is the appraisal (with which I concur) that the units of perception (and of cognition at large) are not atomistic sensory inputs but higher-order meaningful relations. This insight is actually in line with earlier observations by Miller (1956) and his notion of the "chunk." Essentially, Miller's point is that semantic considerations are pivotal in both perception and memory, and that meaning can drastically enhance psychological performance.

A parameter that is especially salient in the writings of Gibson and his followers in the school of ecological psychology is that of time (see Shaw & Pittenger 1978; Turvey et al. 1981). Gibson pointed out that the unit of psychological time, the event, is a rich, higher-order structure, defined by the meaning invested in the organism's activities in the world (Gibson 1960; 1966b; 1975). Shaw and Todd (1980) presented a formal argument showing that the more one takes into consideration the history of a cognitive agent, the less is there a need to postulate underlying representations by way of accounting for this agent's behavior. Eventually, when the entire history of the agent is known, there is no need to postulate any representation (for further development, see Shanon 1993, p. 187).

Interestingly, the points on which the Jamesian and Gibsonian critiques are based are also encountered in the target article. P&V repeatedly show that experimental data are subject to alternative accounts, that tasks claimed to be achieved unconsciously actually involve brief moments of consciousness, and that the power of the cognitive system may be highly increased by enlarging the scope of what are taken to be the units of perception and cognitive processing (in particular, note their "trill" example), a feat involving the appreciation of both meaning and higher-order relations (such as symmetry). In fact, the elemental structures that are perceived can be larger than that.

The two masters I have mentioned were both psychologists and their arguments bear great similarity to ones found in the target article. In conjunction with the mentalist tradition (in the authors' sense) and the dispensing of underlying representations, I shall also mention Merleau-Ponty, a philosopher with direct expertise in experimental psychology. In his *Phenomenology of Perception,* Merleau-Ponty (1962) identifies the domain of the psychological with the domain of the conscious, and argues against the postulation of underlying mental representations. Like James, he argues that between the level of the brain and that of conscious phenomenology, no additional level of covert representations is needed. The postulation of such a level is, of course, the cornerstone of the modern representational-computational paradigm in cognitive science (see, in particular, Fodor 1975). Merleau-Ponty criticizes the simplistic notion of atomistic sensory stimulus and highlights the intrinsic temporality of cognition. Because Merleau-Ponty's arguments are intricate, I shall not belabor them further here.

## Hemineglect, extinction, and the importance of conscious processing

Eric Siéroff

*Laboratoire de Psychologie Expérimentale, Université René Descartes (Paris V) et CNRS, Centre Henri Piéron, 92774, Boulogne-Billancourt, France.*
**sieroff@psycho.univ-paris5.fr**

**Abstract:** Neuropsychological studies on hemineglect and extinction show that "neglected" or "extinguished" stimuli can access a semantic level. However, processing of these stimuli is usually not accomplished at the same level as non-neglected stimuli. These data are compatible with Perruchet & Vinter's hypothesis of the importance of consciousness in the construction of representations and knowledge.

Since Jackson (1876/1932), cerebral lesions have been known to impair propositional and explicit behaviour, sometimes leaving simpler automatic and implicit behaviour intact. In section 8.3 of the target article, Perruchet & Vinter (P&V) question the theoretical pertinence of neuropsychological data on blindsight that show that considerable processing of nonconsciously perceived stimuli is possible. Their arguments are based mainly on the fact that performance is usually degraded. Another syndrome, hemineglect, and some of its clinical signs, for example, visual extinction, have been widely studied to show that neglected stimuli can nonetheless be semantically processed. However, it can be shown that such high-level processing of "extinguished" or neglected stimuli is actually far from being equivalent to the processing of stimuli located on the non-neglected side of space. Consequently, the existing data on implicit processing in hemineglect are compatible with P&V's claim on the importance of consciousness in the construction of representations and knowledge.

Visual extinction is a common phenomenon in neuropsychology, frequently related to the highly devastating disorder of hemineglect (Friedland & Weinstein 1977): Patients can detect and respond to a stimulus presented on the side opposite to their lesion when presented alone, but become unable to do so when another stimulus is presented on the other side, ipsilateral to the lesion. The phenomenon has been related to a bias in orienting of spatial attention (Posner et al. 1984). Volpe et al. (1979) were the first to show some residual processing of the "extinguished" stimuli in four right parietal patients, who presented a quasi-total left extinction for words. When asked to compare the left-sided word with the right-sided word, patients first claimed that no word was present on the left, but they responded above chance level when forced to guess. McGlinchey-Berroth et al. (1996) found that a


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

word presented on the "extinguished" side could produce a strong semantic priming effect on a word subsequently presented on the "intact" side of space. Thus, not explicitly identified words can still access a semantic code. Similar conclusions have arisen from studies using the repetition priming effect (Schweinberger & Stief 2001). Most interestingly, Baylis et al. (1993; see also Baylis et al. 2001) have shown, in patients with partial left extinction, that extinction is stronger when the two items (i.e., letters) are the same than when they are different. A possible explanation is that, because of the deficit in orienting of selective attention, the attributes ("quality") of the stimulus located in the "extinguished" side may not be linked with its location (Treisman 1998): The stimulus is reported only when its attributes are different than the stimulus located on the "good" side of space. A lack of attention does not preclude some high-level processing, but that usually does not allow the integration of all the attributes of the stimulus, therefore explicit identification.

Although patients with left extinction do not explicitly identify the left word in a pair of words, and do not identify the left part of a nonword, they frequently identify the left part of a single familiar word covering the same visual angle as the pair of words (Siéroff & Michel 1987; Siéroff et al. 1988). Single words could be better identified because some lexical processing occurs on the whole letter string, allowing a redistribution of attention over the entire letter string, leading to explicit identification (Brunn & Farah 1991; Siéroff 1991). Presumably, such a redistribution of attention over the entire display is difficult in these patients when the material is complex or new (two letter strings, one long nonword), but is easier when lexical activation occurs. An anecdote about a patient with strong left hemineglect suggests the involvement of consciousness in this phenomenon. I presented words and nonwords of different lengths on individual cards to this patient, who presented overall better explicit identification with words than with nonwords (for which he had a strong tendency to ignore the letters located on the left side). When presented with the longer word of the list: "témoignage" (testimony), he looked at the word and, after two seconds, gave me the card back, saying, "it says 'ignage' (not a word), and it could be 'témoignage' if the rest was present." When asked to check, he persisted in saying "no, only ignage is on the card." Because the word "témoignage" (oi = /wa/) is very difficult to guess on the basis of the right part, "ignage" (i = /i/), it can be hypothesized that the letters on the left side ("témo") activated a lexical unit with the rest of the display. However, the patient's attention was still directed toward the right side of the display, maybe because of the word length, and only this part could be consciously identified: Once again, processing of the neglected side did not reach the same level as the other side.

The residual abilities of the patients on their neglected side should not mask the dramatic clinical consequences of hemineglect. Even if it can be demonstrated that some high-level processing can occur on "extinguished" or neglected material, this processing remains highly degraded, and patients' behaviour is strongly oriented to the other side. Consequently, they cannot accomplish by themselves most spatial tasks in everyday life, such as reading a text, orienting in a new environment, dressing, shaving, eating, and the like. It takes time for some of them to develop strategies to orient attention and conscious processing toward the neglected side: Frequently, they need to verbalize the need to look at the neglected side. For example, in a search task, they ignore targets on the neglected side and, after a while, may say, "Well, if I do not find the target, it could be that it is located on my bad side." This explicit statement, which can be reinforced by rehabilitation, can somewhat compensate for the deficit in orienting of attention. In conclusion, even if hemineglect and extinction data show that some processing can occur without attention and consciousness, attentional and conscious processing appears to be crucial for an oriented and adapted response and behaviour.

## The emergence of consciousness: BUC versus SOC

Ron Sun

*CECS Department, University of Missouri-Columbia, Columbia, MO 65211.*
**rsun@cecs.missouri.edu** **http://www.cecs.missouri.edu/~rsun**

**Abstract:** This commentary suggests that there are two distinct types of interacting cognitive processes. Conscious processes emerge from unconscious processes. The key problem of SOC is that it uses an overly narrow notion of the "cognitive unconscious" to show that the "cognitive unconscious" is not necessary. Yet, it has little to say about the roles of conscious and unconscious processes in general.

Like Perruchet & Vinter (P&V), I would suggest that consciousness self-organizes and emerges. However, anything beyond that is where our agreement ends. The key questions are: Where does consciousness emerge from? Is there anything substantial beneath consciousness? What is there beneath consciousness?

In contrast to P&V's self-organizing consciousness (SOC), the concept of bottom-up consciousness (BUC), which has been developed for some time now (see Sun 1999; 2000; 2001; 2002; Sun et al. 2001), insists on the duality of cognition. In BUC, it is suggested that both conscious and unconscious processes exist in cognition. The essential difference between them lies in representation: Whereas one type of cognitive process uses explicit and accessible representation, the other type uses implicit and relatively inaccessible representation. This representational difference might explain their phenomenological difference (Sun 1999; 2001). In other words, the phenomenological difference might be reduced to the representational difference. The two types of cognitive processes not only coexist, but also interact in complex and multifaceted ways. One of the ways is that conscious percepts, concepts, rules, knowledge, and so on, emerge from unconscious processes in a bottom-up fashion. It has been suggested that this is the essential way by which consciousness emerges (Sun 2002). This view has been substantiated in a computational cognitive model named CLARION (Sun et al. 2001), which simulated a range of psychological data (Sun 2002).

Let us now turn to SOC. In a nutshell, P&V believe in the nonexistence of the "cognitive unconscious" – a vague term that has been used to mean quite a few different things in the past. The authors went to great lengths in defending this thesis. They marshaled a large amount of empirical findings and interpretations (their own and others') in the attempt to shore up their argument against the cognitive unconscious. It is certainly admirable to be this comprehensive. However, the sheer amount of information in this target article makes it rather difficult (even impossible, given the space limitation for commentaries) to dissect their argument in a comprehensive fashion. So, I will focus on only a few key points concerning SOC (vis-à-vis BUC).

With due respect for P&V's huge undertaking, I found that their overall argument was conceptually rather unclear. Intentionally or unintentionally, they greatly broadened the definition of the conscious, and, at the same time, defined the "cognitive unconscious" in an overly narrow way – as a sort of "unconscious rule abstraction" (with some sort of "symbolic" representation) – thus reducing it. Given this definition, it became an easy task to shoot the idea down and declare an easy victory. But, unfortunately, this apparent victory has nothing to do with the cognitive unconscious in general (Sun 2002).

For example, it has nothing to say about whether an instance-based process is conscious or unconscious, or whether an associative learning mechanism is conscious or unconscious, both of which they relied on heavily in the target article. If we adopt the view that an instance-based process or an associative learning process can be (either partially or wholly) unconscious, then their case against the "cognitive unconscious" immediately falls apart. Such an argument for the unconscious is not unheard of (see, e.g., Sloman 1996; Smolensky 1988; Sun 1996; 1999). Alternatively, the



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

authors might argue that instance-based processes and associative learning processes are not "cognitive." But then, what is cognitive? The notion of the "cognitive unconscious" becomes rather uninterestingly narrow, bordering on being empty. In contrast, BUC provides a much richer notion of cognition, and thus of the cognitive unconscious (although this term was not actually used). This richer notion makes detailed mechanistic descriptions of the emergence of consciousness possible (see Sun 2002).

On a related note, the representation/computation tradeoff that the authors cited, and the implied dichotomy of representation versus computation, are, in my opinion, completely orthogonal to the dichotomy of the conscious versus the unconscious. That is, a representation can be either consciously accessible or inaccessible; likewise, a computation can be either conscious or unconscious (Sun 1996). This should be self-evident (if not, lengthy arguments can always be found in various places, for example, Smolensky 1988; Sun 1996; 2000; and various references cited therein). This view is the key to BUC, as mentioned earlier. Yet, the two orthogonal dichotomies were entangled with each other throughout P&V's target article. There are many such confusions in the article.

I can see how the authors came up with this sweeping theory about consciousness – as a result of their experimental work in artificial grammar learning and in other tasks. They have demonstrated some interesting findings that supposedly implicit learning of artificial grammars may be accounted for by "conscious" fragmentary knowledge. But, in my view, this is not a sufficient reason to reject altogether the existence of unconscious processes in cognition, not even with all the similar findings in other domains. We should be careful not to allow limited results from our favorite paradigm to shape our entire world view. Overgeneralizing useful, but limited, results is methodologically harmful, and something we should guard against.

ACKNOWLEDGMENT

This work is supported in part by ARI contract DASW01-00-K-0012.

## Trading automatic/nonautomatic for unconscious/conscious

Joseph Tzelgov

*Ben-Gurion University, Beer-Sheva, 84105, Israel.*
**tzelgov@bgumail.bgu.ac.il** **http://www.bgu.ac.il/~tzelgov/**

**Abstract:** In this commentary I show that the SOC framework implies automaticity of both the materialization of phenomenological conscious experience and the application of the primitives resulting from the emergence of consciousness. In addition, SOC implies that cognition refers to conscious experience. Consequently, I propose automatic/nonautomatic instead of unconscious/conscious as the basic contrast characterizing human cognition.

The main purpose of the work of Perruchet & Vinter (P&V) is to show that there is no need to assume cognitive unconsciousness to explain most phenomena, such as unconscious semantic priming, implicit learning, and automaticity. These can be explained by an alternative, coined self-organizing consciousness (SOC), which assumes conscious representations generated by unconscious processes. The term "representation" as used by the authors refers to "a *mental* event that assumes the function of some meaningful component of the represented world" (see sect. 1.3.2, my emphasis). Representations may hold evocative relations to other representations (they may evoke each other), which are acquired via simple associative learning. This learning process explains the self-organizing aspect of SOC.

The model leads to the conclusion that our mental life consists of conscious representations and, in contrast to Kihlstrom (1987), that there is no unconscious mental life. This does not mean that the information included in the representations evolves from nothing at this very moment. This information is stored in the relevant neural structures. So, one implication (which I agree to) of the model is that cognition starts when stored information changes into representations. It follows that psychology, as a scientific discipline, also investigates phenomena (i.e., inhibition of return; see Posner & Cohen 1984) involved in the generation of these representations or that affect them. These phenomena that interface representations and the neural activity involved in their generation are unconscious in the same sense as any other biological processes.

SOC focuses on evocative events. The evocative events are the proposed alternative to the notion of cognitive unconsciousness. These events correspond to what is done automatically, or in other words, what is not intentionally monitored (Bargh 1992; Tzelgov 1997a). Mental representations can become arguments of deliberative events (sect. 3.1.1), the second basic type of mental events (Dulany 1991; 1997). The deliberative events are the subject of various intentional operations (computations) that are consciously monitored such as rule abstraction, analysis, reasoning, inference, or reading by novices. Thus, deliberative events reflect nonautomatic (or consciously monitored) processing. Most of our mental life consists of evocative episodes (Tzelgov 1997b) and at any single moment our mental life consists of one or more evocative episodes corresponding to various aspects of the world around (the scenes we see, the voices we hear, etc.), and sometimes also to a deliberative event, which corresponds to what we are intentionally thinking of or are doing. This means that, as I suggested elsewhere (Tzelgov 1997b), most of the time we act in a way that is "automatic but conscious."

If all our mental life is conscious, as implied by SOC, cognitive scientists should seriously consider trading "automatic versus nonautomatic" for "conscious versus unconscious" in explaining cognitive processing. The notion of automaticity used here departs from the way the term was used originally by Schneider and Shiffrin (1977) or Hasher and Zacks (1979). It is based on using processing without monitoring as the defining feature, and is best diagnosed by Stroop-like phenomena when it is not part of the task requirement (see Tzelgov 1997a). It views automatic processing as the default mode of processing (Besner et al. 1997; Tzelgov et al. 2000) of specific stimuli in a specific context. The growing of consciousness caused by the emergence of new functional primitives, as combinations of old ones that are constantly within single attentional focus, reflects automatization in this sense and may be seen as redefining the default mode of processing for a set of stimuli. This process is based on automatic learning that does not require intentional orientation toward it (see sect. 2.2.1).

Therefore, in my view, SOC implies that materialization of phenomenological conscious experience (1) is the result of the automatic emergence of new primitives, and (2) reflects automatic processing of these primitives. In line with Logan (1988), SOC proposes to view automatization as a qualitative change – switching from algorithms to single-step memory retrieval. This is easy to accept if *algorithm* is used as synonymous with *intentionally applied* multistep logical operations. It still has to be explained how stored *evocative* representations can explain *automatic* processing that at least appears to include multistep mental activity. Consider once again the issue of transitivity discussed by P&V in section 7.1.1. The direct read-out interpretation makes perfect sense when the participants are asked to intentionally perform a task conforming to applying transitivity.

Such an interpretation is less trivial when applying transitivity is not part of the task requirements. Tzelgov et al. (2000) trained subjects in magnitude comparisons by presenting them with pairs of arbitrary figures and asking them to decide which represents a larger magnitude. The figures corresponded to the numbers 1–9, and only pairs corresponding to adjacent numbers were included in training. After training, presenting pairs of *nonadjacent* "artificial numbers" for *physical comparisons* resulted in a size congruity effect. Thus, assuming that Δ and Ξ are two arbitrary fig-


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

ures corresponding to 3 and 5, responses were faster to the pair Δ Ξ than to the pair Δ Ξ. Explaining these findings requires assuming that the training resulted in generating a representation that not only aligns all the arbitrary numbers on a single number line but also includes as a default (or automatically results in) the activation of their relative magnitudes whenever a pair of such numbers is presented (at least for some kind of size comparisons). As another example, it has been shown (Tzelgov et al. 1996) that the Stroop effect can be obtained even when color names are written in a script that they never appeared in (e.g., by presenting Hebrew color names written in English to Hebrew–English bilinguals). This can be explained in terms of strong phonological theory (Frost 1998) if one assumes that the representations of letters include their mapping to the corresponding phonemes. Thus, to be considered as an alternative for the traditional view, SOC has to include mechanisms that mimic intentionally applied algorithms.

## Could phenomenal consciousness function as a cognitive unconscious?

Max Velmans

*Department of Psychology, Goldsmiths College, University of London, London, SE14 6NW, United Kingdom.* **m.velmans@gold.ac.uk**
**http://www.goldsmiths.ac.uk/departments/psychology/velmans.htm**

**Abstract:** Evidence for unconscious semantic representation suggests that a cognitive unconscious exists. Phenomenal consciousness cannot easily be shown to deal with complex cognitive operations such as those involved in language translation and creativity. A self-organising phenomenal consciousness that controls brain functions also runs into mind/body problems (well recognised in the consciousness studies literature) that Perruchet & Vinter must address.

The existence of a complex cognitive unconscious is commonly thought to be one of the major discoveries of twentieth-century psychology, and the evidence for it is extensive (see, e.g., Dixon 1981; Kihlstrom 1987; 1996; Velmans 1991). Given this, Perruchet & Vinter's (P&V's) vigorously argued case that *none* of this evidence is convincing and that *all* of this evidence can be explained in terms of the operation of a self-organising phenomenal consciousness is certainly bold, and deserves to be considered seriously. I found much of their case well argued, but parts of it unconvincing – for example, their claim that no unconscious semantic representations exist. Not all the evidence for this is as easily dismissed as they suggest, for example, the evidence found by Groeger (1984a; 1984b; 1988) of preconscious semantic analysis in nonselected channels. Groeger demonstrated that words in a nonattended ear could bias the meanings of attended-to words, and crucially, he found that the effects of nonattended words were different if they were *above threshold* (consciously detectable) versus *below threshold.*

For example, in one experiment subjects were asked to complete the sentence "She looked ___ in her new coat" with one of two completion words, "smug" or "cosy." Simultaneous with the attended sentence the word "snug" was presented to the nonselected ear (1) above threshold, or (2) below it. With "snug" presented above threshold, subjects tended to choose "smug," which could be explained by subjects becoming momentarily aware of the physical form of the cue. With "snug" presented below threshold, subjects tended to choose "cosy" indicating semantic analysis of the cue word without accompanying awareness. In this case, Holender's (1986) appeal to attentional switching will not work. In other cases, Perruchet & Vinter's alternative explanations do not seem sufficient to accommodate well-accepted evidence – for example, the evidence that a genuinely creative solution to a problem can suddenly appear in consciousness after sleeping on it (incubation). Can this really be explained in terms of selective forgetting of inappropriate associations (sect. 7.1.2)? If creativity were solely a matter of retrieving appropriate, *existing* associations, what about creativity could possibly be *creative*? And how can one make sense of the extensive literature on the existence and operation of long-term semantic memory, a store of global knowledge about the world commonly thought necessary to make *sense* of language as opposed to merely segmenting it into words. Consider, for example, what is required to sort the following 16 sentences into similar meaning pairs:

(1) A nod is as good as a wink. (2) An unfortunate experience produces a cautious attitude. (3) Every cloud has a silver lining. (4) Fine feathers make fine birds. (5) Hints are there to be taken. (6) Idealists can be a menace. (7) It is an ill wind that blows no good. (8) Least said, soonest mended. (9) Never count your chickens before they are hatched. (10) Never judge a sausage by its skin. (11) Once bitten, twice shy. (12) Reality imposes its own limitations. (13) Some disagreements are best forgotten. (14) The road to hell is paved with good intentions. (15) There's many a slip twixt cup and lip. (16) You can't make a silk purse out of a sow's ear (from Green 1981, p. 180).

As Green notes, any reasonably intelligent adult will sort these into eight, similar meaning pairs with little difficulty, against odds of over 13 million to one – whereas no machine currently on the stocks, using a general programme, would do better than chance. Why not? In these pairs, similar ideas are conveyed by sentences composed of entirely different words embedded in different surface forms (compare, for example, (f) and (n)) and their meaning cannot be understood without a global understanding of the physical and social world. At any given moment, this store of world knowledge will be largely unconscious. Without access to an equivalent knowledge store, this task is difficult for machine translation as it involves far more than the manipulation of individual word semantics according to syntactic rules. It would be even more difficult, I suspect, for PARSER, a programme that does not even incorporate semantic and syntactic rules.

It should be evident that P&V's task is Herculean, and I will leave it to others to comment on this aspect of their thesis in more detail. I am more interested in their alternative thesis – that the operations of mind conventionally attributed to a cognitive unconscious are in fact carried out by a self-organising phenomenal consciousness. What could this possibly mean? How could a *subjective experience* influence the activities of brain? In my own analysis of how phenomenal consciousness relates to human information processing (Velmans 1991), I pointed out that the term *conscious process* is ambiguous. A process might be conscious: (1) in the sense that one is conscious *of* the process, (2) in the sense that the operation of the process is *accompanied* by consciousness (of its *results*), and (3) in the sense that consciousness *enters into* or *causally influences* the process.

We do not have introspective access to how the preconscious cognitive processes that enable thinking produce individual, conscious thoughts in the form of inner speech. However, the content of such thoughts and the sequence in which they appear does give some insight into the way the cognitive processes (of which they are manifestations) operate over time in problem solving, thinking, planning, and so on (such insights were incorporated into the broad design principles of the "General Problem Solver," an AI system developed by Newelet al. 1960). Consequently, such cognitive processes are partly conscious in sense (1), but only insofar as their detailed operation is made explicit in conscious thoughts, thereby becoming accessible to introspection. Many psychological processes are conscious in sense (2) but not in sense (1) – that is, we are not conscious of how the processes operate, but we are conscious of their *results.* This applies to perception in all sense modalities. When consciously reading this sentence, for example, you become aware of the printed text on the page, accompanied, perhaps, by inner speech (phonemic imagery) and a feeling of understanding (or not), but you have no introspective access to the processes which enable you to read. Nor does one have introspective access to the *details* of most other forms of cognitive functioning, for example, to the detailed operations that enable "con-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

scious" learning, remembering, and engaging in conversations with others.

Crucially, having an experience that gives some introspective access to a given process, or having the results of that process manifest in an experience, says nothing about whether that experience *carries out* that process (in the sense that Perruchet & Vinter claim). That is, whether a process is *conscious* in sense (1) or (2) needs to be distinguished from whether it is conscious in sense (3). Indeed, it is not easy to envisage how the experience that makes a process conscious in sense (1) or (2) *could* make it conscious in sense (3). Consciousness *of* a physical process does not make consciousness responsible for the operation of that process (watching a kettle does not determine when it comes to a boil). So, how could consciousness *of* a mental process carry out the functions of that process? Alternatively, if conscious experience *results* from a mental process, it arrives *too late* to carry out the functions of that process. Conscious experiences might of course influence neural processing that *follows* its emergence. But this too is problematic. Most scientists would agree that phenomenal consciousness has physical correlates that can in principle be viewed from a third-person perspective. Most scientists would also agree that, viewed from a third-person perspective, the physical world (including the brain) is causally closed. If so, any role attributed to phenomenal consciousness will already have been fulfilled by its physical correlates (this is sometimes referred to as the problem of overdetermination). If so, what remains for phenomenal consciousness to do? There are many additional, deeper problems (extensively discussed in Velmans 2000; 2002a; 2002b). Nevertheless it would be useful to learn how P&V would deal with these preliminaries, to get their theory about a self-organising phenomenal consciousness off the ground.

## Language heterogeneity and self-organizing consciousness

William S.-Y. Wang and Jinyun Ke

*Department of Electronic Engineering, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China.* **eewsyw@cityu.edu.hk**
**jyke@ee.cityu.edu.hk** **http://www.ee.cityu.edu.hk/~wsyw/**

**Abstract:** While the current generative paradigm in linguistics leans heavily toward computation, investigations on conscious representations are much welcome. The SOC model examines the acquisition of complex representations in individuals. We note that heterogeneity of representation in populations is a central issue that must be addressed as well. In addition to the self-organizing processes proposed for the individual, interactions among individuals must be incorporated in any comprehensive account of language.

Perruchet & Vinter (P&V) have examined the roles of representation and computation within the mentalistic framework suggested by Dulany. While the literature in linguistics over the last several decades focuses overwhelmingly on computation, that is, on how sentences are generated or computed by rules that operate on what the authors call unconscious representations, earlier deliberations on this issue have not always been so one-sided. For example, in attempting to weigh rules against pre-stored language constructions, Dwight Bolinger asked in 1961:

> Is grammar something where speakers "produce" (i.e., originate) constructions, or where they "reach for" them, from a pre-established inventory, when the occasion presents itself? If the latter, then constructions are not produced one from another or from a stock of abstract components, but filed side by side, and their interrelationships are not derivative but mnemonic.

The early 1960s was the period when American linguistics switched paradigms from structuralism to generativism and shifted attention from phonology/morphology to syntax. In efforts to formalize syntactic dependencies, say, between X and Y, researchers were fascinated by the fact that X and Y can be separated indefinitely far from each other by intruding constructions, that the intruding constructions themselves may contain discontinuous constructions, and that these dependencies may even cross over each other, such as in constructions built on the word *respectively.* Much of linguistic theorizing over the past decades was shaped by these initial efforts of formalization, and sometimes highly abstract computational machinery has been invented to explain rather marginal if not illusory linguistic phenomena.

In contrast with the generative paradigm, which assumes homogeneity in the rules speakers use, many linguists are now paying increasing attention to the fact that speakers are tremendously diverse and varied in their language abilities and behaviors. An early statement of this realization is that of John R. Ross (1979) in his paper "Where's English?":

> The view of language that seems most plausible to me is that the sentences of a language are points in an *n*-space. An idiolect is a vector in this *n*-space. And each speaker's vector, or path, through the space will, I expect, be as individual as his or her face – a linguistic fingerprint.

Individuals certainly differ greatly in their preferred constructions, especially those prefabricated ones, be they cliches, epigrams, pause fillers, or whatever. English is actually a relatively poor language for studying these prefabs because of its relative youth; and its international status has bleached it of many language-specific characteristics (Wang 1991). More local languages with longer independent histories are much richer in such constructions, which are clearly more like conscious representations rather than unconscious representations.

At any rate, linguists who are impressed with heterogeneity and variability in language, and who resonate with the remarks quoted here from Ross (1979) and Bolinger (1961), would welcome P&V's emphasis on conscious representations. P&V reject "the notions of unconscious rule abstraction, computation, analysis, reasoning, and inference. Because unconscious representations have no other function than to enter into these activities, eliminating the possibility of these activities actually makes the overall notion of unconscious representation objectless" (target article, sect. 1.3.1). While this statement may turn out to be too categorical and extreme, it is important to learn to what extent it can be pushed.

There is currently an intense debate on whether there is some mechanism that is specific and exclusive to language, and whether this mechanism is genetically determined (see, e.g., Schoenemann & Wang 1996). The notion that language evolves to be adaptive to human's cognitive capacity has been gaining more and more attention of late (Christiansen 1994; Deacon 1997); it is mentioned in the target article as well. To investigate how language has evolved calls for research with a population perspective, in addition to the viewpoint of the individual that the model of P&V primarily deals with. With models from a population perspective, we see that many complex structures in language are actually the result of long-term evolution; no powerful language-specific processing module in the individual brain is necessary. There is a class of models that also makes important use of *self-organization* in pursuing this line of thought. They simulate how language evolves in a population, such as the model for the emergence and evolution of lexicon (Ke et al. 2002; Steels 1997), the emergence of sound systems (de Boer 2001), and the like. In these models, linguistic structures emerge from rudimentary states through interactions among individuals who are endowed with some basic perception and production constraints. Only very simple mechanisms, such as imitation, are required. We see this class of models as complementary to the SOC model in studying the two scales of language evolution, specifically, ontogenetic and phylogenetic.

The SOC model is a plausible framework in addressing the ontogenetic development of language in individuals. In the PARSER model, the primitive units are extracted and accumulated from a given complex text containing underlying structures. In the latter class of models, primitives are largely given, for example, the phones in phonology models and the words in lexicon models.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Structures, such as a relatively stable vowel system or a lexicon that is shared by the individuals in the population, emerge by the diffusion of these primitives. These models are more applicable for phylogenetic emergence and language change. The self-organization mechanism that takes effect through the interactions between language users, leads to continuous change, which is a hallmark of language.

Lastly, we note that the (quick) forgetting constraint is a key feature in the PARSER model, and that it presents an important parallel to the selection mechanism in evolution theory. In the PARSER model, those chunks that do not repeat themselves enough within a given time span will be forgotten and therefore cannot become primitives. This feature roughly corresponds to the "survival of the fittest" mechanism in Darwinian evolution theory, which says those individuals not adaptive to the environment, that is, having a bad fitness, will have less chance to pass their genes to the next generation. Therefore the genes that happen to exist only in unfit individuals cannot be sustained and will be lost from the gene pool. We can see these effects in many simulation models that implement these principles, such as the Genetic Algorithm model used in simulating the emergence of vocabulary (Ke et al. 2002) and the population dynamics model in simulating the evolution of universal grammars (Nowak et al. 2001).

## Contrasts and dissociations suggest qualitative differences between conscious and unconscious processes

Gezinus Wolters[a] and R. Hans Phaf[b]

[a]*Department of Psychology, Leiden University, Leiden, 2300, The Netherlands;* [b]*Psychonomics Department, University of Amsterdam, Amsterdam, 1018, The Netherlands.* **wolters@fsw.leidenuniv.nl** **phaf@psy.uva.nl**

**Abstract:** The authors reject a computationally powerful unconscious. Instead, they suggest that simple unconscious processes give rise to complex conscious representations. We discuss evidence showing contrastive effects of conscious and unconscious processes, suggesting a distinction between these types of processes. In our view, conscious processes often serve to correct or control negative consequences of relatively simple unconscious processes.

In an admirably radical attempt to clarify matters regarding conscious-unconscious distinctions, Perruchet & Vinter (P&V) settle the issue, almost by definition, by proposing that all processes are unconscious, whereas all representations should be conscious (i.e., a mentalistic framework; Dulany 1991; 1997). Alternatively, it can be argued, as has been put forward by critics of the computer metaphor, that there is no real distinction between process and representation (or between central processor and memory system, as in a computer). In this view, the representation consists of the collection of processes it affords.

Although the range of empirical phenomena covered by P&V is extremely broad, it cannot be avoided – in view of the generality of their claim – that they missed some relevant findings. We agree that evidence showing "complex" (e.g., semantic) processing in the absence of conscious detection (i.e., an indirect-without-direct effect; Reingold & Merikle 1988) cannot refute the mentalistic framework. With the indirect-without-direct pattern of results, it can always be argued that the conscious (i.e., direct) effects have not been exhaustively excluded and that some conscious effects have "leaked through" to unconscious conditions. However, empirical contrasts or dissociations between conscious and unconscious (or less conscious) conditions (Baars & McGovern 1993; Merikle 1992) are much harder to reconcile with the mentalistic framework. If qualitative differences are obtained as a function of relative level of consciousness, this can hardly be blamed on the sudden transition from a conscious representation to an unconscious process. Because the same dependent variable is addressed, effects must be accounted for either only in terms of processes or representations. We will briefly discuss three examples of experimental contrasts that, in our opinion, challenge the mentalistic framework.

First, the treatment of implicit and explicit memory performance by P&V (sect. 8.1) implies that they view both types of tests as essentially measuring the same thing, but that the tests differ in sensitivity. A large body of evidence (Richardson-Klavehn & Bjork 1988; Roediger & McDermott 1993) showing contrastive effects as a function of the same manipulation, however, suggests otherwise. Of course, it can be argued by P&V that different aspects of the conscious representation were addressed because the format of the two types of tests generally differs. A purer form of measuring both types of performance, where test format does not differ, may be found, however, in the Process Dissociation Procedure (PDP, e.g., Jacoby et al. 1993). Using this procedure, Jacoby et al. (1993, Experiments 1, 2) replicated Eich's (1984) finding of a dissociation between implicit and explicit memory performance as a function of level of attention at study. Interestingly, with small variations in consciousness, as under different levels of anaesthesia, it is also primarily (the already low level of) explicit memory that decreases with depth of anaesthesia (Phaf 2001). The crucial finding is that, when measured with PDP, explicit memory decreases as a function of consciousness at a very much faster rate than implicit memory. Even when implicit memory is also found to be reduced by attentional manipulations (Wolters & Prinsen 1997), it remains on the mentalistic framework to explain why the rate differs so markedly from explicit memory.

Second, P&V discuss the apparent indirect-without-direct effect obtained by Cheesman and Merikle (1986) in a Stroop task (sect. 8.2.3). Also, the contrast between conscious and less conscious conditions, which is much more interesting in the present context, is ignored. Merikle and Joordens (1997) presented an improved form of this contrast. They showed that by decreasing consciousness (because of either divided attention or impoverished presentation) the reversed Stroop effect induced by unequal proportions of congruent and incongruent trials returned to the normal Stroop effect. Such a finding presents a problem for the mentalistic framework because it implies that processes change qualitatively when consciousness is reduced and thus argues in favour of both unconscious and conscious processes.

Third, a pattern of results that is particularly hard to reconcile with consciousness leaking through to unconscious conditions occurs when effects become stronger with a reduction of consciousness. Although not exclusive to affective processing (e.g., De Fockert et al. 2001), it seems to be predominant in this field (Murphy & Zajonc 1993; Rotteveel et al. 2001). Priming by faces with emotional expressions on evaluations of Chinese ideographs is, generally, larger in suboptimal (i.e., less conscious) than in optimal (i.e., fully conscious) prime presentation conditions and may even be reversed to incongruent priming in the latter conditions. Again, these results argue in favour of qualitatively different conscious and unconscious processes, which may be captured very well by the neurobiological dual-route model of LeDoux 1996 (for connectionist simulations of affective priming, see Capalbo et al. 2000). The dual-route architecture was initially developed to account for fear conditioning in both humans and animals and was based on findings that conditioning could take place through both routes separately. Contrary to what is argued by P&V, it is hard to imagine that conditioning through the direct, fully subcortical route always requires conscious cognition both in humans and animals. Indeed, learning occurs, for example, in newborns and in aplysia (Kandel 1991), who presumably do not dispose of conscious representations.

The acceptance of both conscious and unconscious processes does not necessarily entail a powerful *cognitive* unconscious, as is assumed sometimes in computational theories of mind. In fact, the main qualitative difference may be that unconscious processes are much less powerful than conscious processes (see Greenwald



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

1992), as can indeed be seen in the three examples described earlier. More than as a specific model for fear, LeDoux's dual-processing architecture may serve as a general framework for viewing the relation between conscious and unconscious processing (Phaf & Wolters 1997). Action tendencies can be triggered in parallel by fast and direct (unconscious) processes and by slower and indirect (sometimes conscious) processes. Particularly when direct processing is mandatory, conscious processing can serve to modulate or correct negative consequences of these, unconsciously evoked, action tendencies.

## The self-organizing consciousness entails additional intervening subsystems

Takashi Yamauchi

*Department of Psychology, Texas A&M University, College Station, TX 77843.* **tya@psyc.tamu.edu**

**Abstract:** The self-organizing consciousness (SOC) is not sufficient to account for young children's ability to acquire complex rules and word-object mappings. First, the attention-association cycles suggested by the SOC are unlikely to happen because recurrence of particular stimulus properties usually disengages the attention of an observer. Second, "primitive processors" preinstalled in the system make the SOC unnecessarily complex.

From about 18 months to 6 years of age, children acquire more than 15,000 new words. Putting it differently, children learn 10 new words every day during this 4.5-year period.

To form associations between words and objects, and to maintain these associations in different contexts, children need to experience the same words and objects many times in close temporal and spatial proximity. If the SOC (self-organizing consciousness) is the only system that mediates word-object linkages, then children have to discover at least 10 correct associations everyday, 70 in one week, 300 in one month, and 3,650 associations in one year. If about five self-organizing cycles between attention and association are needed to acquire associative chunks, then 50 cycles a day, 350 associative revisions a week, 1,500 improvements in a month, and 18,250 iterations in a year should be dispatched consciously by a child during this 4.5-year period. Clearly, this feat is close to impossible for adults. If the same SOC mechanism is at work in adults and young children, why is this enormous chunking process relatively easy for children?

The kernel of the SOC is progressive cycles of attention and associative learning. Attention guides associations because whatever in the attentional focus at a given point gets associated. The progression of attention-association cycles iteratively refines and revises old chunks. Whimsical elements of a stimulus will be dropped out from the chunking process because these elements are unlikely to recur in different contexts, just as noise is canceled out, and the true mean of a population is gleaned from many samplings.

For example, by hearing a sentence "look at the dog," infants eventually associate the word "dog" with a moving four-legged, doglike animal after experiencing repeated occurrences of the word "dog" and some related animals. Other erroneous chunks such as "look" and the color of the animal are unlikely to happen repeatedly in different contexts; therefore random chunks die out. Because distal stimuli embody the structure of the world one way or another, the complex representation that is isomorphic to the world itself emerges because of these progressive interactions between attention and associations. In this manner, attention and association play a reciprocal role; attention guides associative learning, and in turn associations guide attention. That is the reason why the system is called "self-organizing consciousness."

However, for the SOC to be truly functional, it requires some other intervening subsystems. Empirical research seems to suggest that simple recurrence of some previously formed chunks is likely to quell attention rather than to arouse attention. Recurrence of a particular stimulus property usually disengages the attention of an observer. Habituation in infants is a case in point. Learning in general reduces the activity of a designated part of the brain. So, if some meaningful property appears repeatedly over different contexts, then attention to that property should be reduced as learning progresses.

If attention (or conscious experience, as Perruchet & Vinter [P&V] call it) still guides associative learning (as suggested by P&V), and if attention helps extract some abstract relational properties of stimuli, then there should be other intervening subsystems, which can maintain attention despite the recurrence of the same properties. Clearly, the SOC by itself is not sufficient to account for the learning ability of young children.

To extract complex stimulus relations, such as the abstract sequence of syllables used by Marcus and his colleagues (Marcus et al. 1999), the SOC makes use of "perceptual primitives," which are tuned to a complex abstract rule. For example, infants learn to distinguish a sequence of syllables such as "wofefe" and "wefewo" even though these syllables never appeared during training. For this to happen, the SOC employs a primitive processor that extracts the abstract rule such as same-different-same versus same-same-different (McClelland & Plaut 1999). This means that, as Marcus (1999) correctly pointed out, the associative mechanism is not really finding the rule "internally," but the rule is given "externally."

To acquire complex rules such as syntax, how many preinstalled primitives are needed? The processor that traces same-different relations alone will not be sufficient. For each different abstract relation, the SOC needs different primitives. Suddenly, the SOC becomes not so parsimonious.

When all factors are equal, it is safe to choose the theory with fewer assumptions. The principle of parsimony is a tool to separate "good theories" from "bad theories" when these theories account for the data equally well. It is a tool to provide guidance, not evidence to support a particular theory. The reason is simple. Because each assumption requires extra testings, fewer assumptions mean a smaller probability that the assumptions go awry. It is a probabilistic guideline (Goodson & Morgan 1976; Marx 1976).

It is important to develop a parsimonious theory. However, whether or not the brain organizes cognitive functions in accordance with the principle of parsimony is an open empirical question.

P&V's article helps us realize, quite convincingly, that much of complex rule-oriented behavior can potentially arise from simple associative learning. At the same time, the target article reminds us how difficult it is to explain every cognitive activity with simple associative learning.

# Authors' Response

## The self-organizing consciousness as an alternative model of the mind

Pierre Perruchet and Annie Vinter

*Université de Bourgogne, LEAD/CNRS, 21000 Dijon, France.*
**pierre.perruchet@u-bourgogne.fr** **annie.vinter@u-bourgogne.fr**
**www.u-bourgogne.fr/LEAD**

**Abstract:** Through the concept of self-organizing consciousness (SOC), we posit that the dynamic of the mind stems from the recurrent interplay between the properties of conscious experiences and the properties of the world, hence making it unnecessary to



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

postulate the existence of an unconscious mental level. In contrast, arguments are provided by commentators for the need for a functional level of organization located between the neural and the conscious. Other commentaries challenge us concerning the ability of our model to account for specific phenomena in the domains of language, reasoning, incubation, and creativity. The possibility of unconscious semantic access and other alleged instances of adapted performance in the absence of any conscious counterpart are also put forth as evidence against our view. Our response emphasizes the fact that opponents to our model often present as factual, theory-free evidence which is in fact nothing more than the postulates underlying the classical computational framework.

Our target article proposed a general view of the mind intended to serve as an alternative to the classical computational framework. We are grateful to the commentators, including those whose commentaries do not appear in this issue, for their insightful remarks. In this response, we first draw an outline of our proposition (sect. R1). The next two sections are intended to specify the meanings we ascribe to key concepts, especially consciousness (sect. R2), processes, and representation (sect. R3). The commentaries that challenge the nature of the model itself (sect. R4), or contest its relevance in some specific areas of application (sects. R5 to R7), are then dealt with. Finally, we address the question of whether, despite the still sketchy state of the SOC concept, it may be thought of as a promising alternative to the classical framework (sect. R8).

## R1. Restating the SOC model

### *R1.1. An initial dissatisfaction with the classical framework*

The classical cognitive framework represents an attempt to understand how human subjects process information, irrespective of whether the processed information is conscious or unconscious. Two major consequences followed from this starting point.

First, this led to the positing of the existence of a so-called "cognitive unconscious." The term has had different resonances for different commentators, from **Velmans**, who claims that "the existence of a complex cognitive unconscious is commonly thought to be one of the major discoveries of twentieth century psychology," to **Sun**, who sees it as "a vague term that has been used to mean quite a few different things in the past." Thus, a rather more detailed examination is in order.

Let us take one example in each of three domains: perception, language, and reasoning. In the perception area, Rock (1984) says that perception is intelligent and goes on:

> I mean to say that it is based on such thoughtlike mental processes as description, inference, and problem solving, although these processes are rapid-fire, unconscious, and nonverbal. . . . "Inferences" implies that certain perceptual properties are computed from given sensory information using unconsciously known rules. For instance, perceived size is inferred from the object's visual angle, its perceived distance, and the law of geometrical optics relating the visual angle to object distance. (Rock 1984; see Searle's comments on this quotation, Searle 1992)

In the domain of language, the past-tense inflection of verbs in English has become a paradigmatic example. According to Pinker, the use of an abstract rule is necessary to account for regular verb forms. He finds evidence for a system that is "more sophisticated than the kinds of 'rules' that are explicitly taught" (Pinker 1991). **Barrouillet & Markovits** provide us with a further example in the reasoning area with Oaksford's and Chater's model:

> Oaksford and Chater (2001) suggested that reasoning is based on probabilities computation. For example, they assume that when solving Wason's selection task, individuals calculate probabilities using Bayes' theorem. It can be inferred that these computations are unconscious, particularly since few individuals can successfully perform these complex calculations when explicitly asked to do so.

These examples (others were provided in the target article, sect. 1.2) are representative of the classical theories in cognitive science. One characterization of these theories is, to borrow **Carlson**'s formulation, that they "postulate states that are essentially *like* conscious mental states but not part of the stream of subjective experience. For example, we know what it is like to hold a rule (consciously) in mind and use it to control our behavior." The classical approach accepts that there are "closely analogous processes that involve similar representations but are unconscious." As illustrated by the examples above, the only difference, except for their unconscious status, is that the rules involved in unconscious computations may be more complex than those we can manage consciously. Of course, in most cases, the postulate of an unconscious device that works as conscious thought does is not explicit. The explicit conjecture is that human information processing works as a computer does. However, because computers themselves have been conceived on the model of conscious thought, the ultimate model for information-processing theories is indeed the conscious mode of functioning.

The second consequence of focusing on information processing without taking the conscious status of the processed information into consideration has been the nearly complete neglect of consciousness. Because there is no principled limit to the power of the cognitive unconscious, the latter has turned out to be able to provide an account for a huge amount of data. When more recently researchers returned to the study of consciousness, consciousness appeared as the supernumerary piece of the jigsaw. Its function has been confined to some more or less epiphenomenal access to the end products of unconscious computations. In all cases, compared to its unconscious counterpart, consciousness is conceived of as notoriously underpowered, its main attributes being limited capacity, seriality, and slowness.

Although our familiarity with the current landscape may undermine our capacity for judgment, it is hard to escape the conclusion that the classical view of mind is singularly awkward and unparsimonious. This view postulates an unconscious device endowed with extrapowerful computational tools, whereas our conscious mental life is condemned, for some mysterious reasons, to rely laboriously on far more limited capacities. Where is the mischievous sprite that prevents us from gaining deliberate access to the marvelous device that cognitive scientists imagine?

### *R1.2. The self-organizing consciousness*

In striking contrast with the classical framework, we start from the idea that a functional approach cannot be carried out without simultaneously addressing the issue of consciousness, and vice versa. Thus we reject both the sugges-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

tion of **Kurthen** of "abandoning the whole question of whether or not a property or process or entity is conscious or unconscious," and **Litman & Reber**'s claim according to which the only valid question concerns whether a problem is solved consciously or unconsciously, irrespective of the particular strategy used to come to the solution.[1]

In the target article, the proposed research objective was to seek an account for the fact that conscious representations are isomorphic to the world. Our purpose was to explain why we generally perceive continuous speech as a meaningful sequence of words and the visual environment as composed of persons and objects, or why we are able to find the solutions to problems. **Rivera-Gaxiola & Silva-Pereyra** and **Manzotti & Sandini** were concerned about the concept of isomorphic representations, and we believe their remarks to be sound. Manzotti & Sandini, for example, noted: "If a spatial relation can be somehow duplicated by some kind of spatial representation, how would be possible to duplicate taste, smell, or other kinds of qualitative conscious contents?" We agree that many aspects of our conscious experiences do not enter into the concept of isomorphic representations but nevertheless need to be accounted for. For example, it is indisputably adaptive to be disgusted by a food that can made us sick, although our conscious experience, in this context, may be thought of neither as isomorphic to the external world nor even as a representation of the world. We now prefer to state our ultimate objective as one of accounting for "adapted conscious experiences" rather than "isomorphic representations." The concept of isomorphic representations certainly concerns a subset of the events of interest but is too narrow to serve as a general research objective. Note that we do not consider this objective as identical to searching for an explanation for overt behavior. As will be detailed in section R8.1, our objective can be viewed as a part of a more general attempt to account for adaptive behavior. However, we also believe that, to a large extent, this objective does not substantially differ from that of most experimental researchers, at least when it is inferred from their practice instead of being drawn from explicit declarations (Note 5 in the target article).

The solution we propose endorses a dynamical perspective. We account for the content of conscious experiences at time $t$ by self-organizational processes acting on the content of conscious experiences at time $t-1$. The change from $t-1$ to $t$ 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 world.[2]

The self-organization of conscious thought was illustrated at length in the target article on the basis of a situation in which people learn to perceive the words of an artificial language through their exposure to an unsegmented speech flow. This situation is specially interesting because both formal and connectionist models have been proposed in the past as ways of discovering the words composing the language. The formal models (e.g., Brent 1996) are based on mathematical algorithms of optimization, whereas connectionist models (e.g., Elman 1990) perform statistical analyses of the input. In both cases, the discovery of words implies computations that learners are not aware of performing and furthermore should be unable to perform if they were asked to do so. In both classes of models, the phenomenal experience of the learner is unrelated to the learning process. Conscious knowledge of the words is thought of as an epiphenomenal access to the end product of unconscious computations.

By contrast, in the model we propose (the computer implementation of which is named PARSER; see Perruchet & Vinter 1998b), the discovery of words stems from the gradual organization of the phenomenal experience of the learner. The model starts from the initial perception of random fragments of speech and shows how the correct perception of words emerges without performing any form of mathematical or statistical computation. Words emerge thanks to a few properties inherent to conscious thought, essentially that (1) any conscious experience embeds a few primitives; (2) these primitives tend to form, by association, a new perceptual unit; and (3) the new unit is either reinforced by repetition or otherwise forgotten. These properties are efficient because of the properties of language. Interestingly, these latter are the very same properties as those exploited in the formal and connectionist models. However, instead of exploiting the properties of language explicitly through appropriate analytic tools, our model exploits these properties in a dynamic, interactive way, thanks to the properties of conscious states.[3]

Our target article was dedicated to showing that what is true for the word segmentation setting is in fact generalizable to all situations of interest to psychologists.

### *R1.3. A mentalist metatheory*

To the best of our knowledge, the proposal that conscious contents self-organize is ours. However, our focus on consciousness – and the consequent mistrust of a cognitive unconscious – is hardly new. As pointed out by **Shanon**, William James was a famous precursor. Interestingly, his position was not because of an ignorance of other views of the mind. Rather, his proposal took the form of a principled argumentation against those who, before him, had stated their belief in the power and sophistication of unconscious thoughts. Although working within different perspectives, Merleau-Ponty (1962), who is also cited by Shanon, and Searle (1992) both argued along similar lines from a philosophical standpoint.

In the target article, we committed ourselves to what is known, after Dulany (1991; 1997), as a mentalistic metatheory. To quote **Carlson**, "the mentalistic position is that no states like conscious mental states (but unconscious) need be invoked to explain psychological phenomena." According to **Dulany**, who outlines the main tenets of mentalism in his commentary, "a mentalistic metatheory leaves no place for a resident humanoid that secretly does what persons do – the cognitive unconscious." Rejection concerns not only unconscious rule abstraction but the notion of unconscious representation as well. The only operations we perform on representations, and moreover the only representations we have, are those that compose our momentary phenomenal experience. Mental life is coextensive with consciousness.

Although we subscribe to the main tenets of **Dulany**'s mentalistic metatheory, it is worth emphasizing that our article was not aimed at demonstrating that a cognitive unconscious does not exist. In fact, proof of nonexistence is beyond the scope of any empirical approach. However, this in no way undermines our intentions because we actually raised a different question. The correct question, we argue, is not: "Does the cognitive unconscious exist?" but instead,


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

"Is the cognitive unconscious necessary?" Our objective was actually to show that the cognitive unconscious is not needed because consciousness is sufficient (sect. 1.4.1). We are aware of the paradox for most readers, because the issue of consciousness is most often framed in the opposite direction. We justify our own approach by the following reasoning. Everyone agrees, we guess, that the cognitive unconscious is (1) a postulate, and (2) subject to no principled functional limitations. By contrast, consciousness is (1) a fact that is observable by all, and (2) subject to significant constraints (limited capacity, seriality, slowness). Asking whether consciousness is sufficient makes sense, we claim, because it amounts to (1) preferring a fact to a postulate, and (2) choosing the less powerful (and hence more parsimonious) option.

## R2. About consciousness

Not surprisingly, our focus on the conscious/unconscious distinction raised many comments and criticisms. Some of them have resulted in substantial additions to the target article, even though we acknowledge from the outset that some issues remain open.

### *R2.1. Consciousness and metaconsciousness*

In the target article, the recurrent use of expressions such as subjective or phenomenal experience, without further specification, was intended to cover a very wide range of events. In need of a definition, we could adopt Searle's (e.g., 1990) famous one:

> By "consciousness" I simply mean those subjective states of sentience or awareness that begin when one awakes in the morning from a dreamless sleep and continue throughout the day until one goes to sleep at night or falls into a coma, or dies, or otherwise becomes, as one would say, "unconscious."

In the target article, we alluded to Dulany's (e.g., 1997) distinction between evocative and deliberative conscious events. **Tzelgov** cogently comments on this distinction. Evocative episodes, which fill up most of our mental life, correspond, he argues, to what is done automatically. The deliberative events correspond to the less frequent occurrences of consciously monitored operations of analysis, rule abstraction, inferences, and planned action. Tzelgov proposes distinguishing between the two categories of events using the automatic versus nonautomatic distinction. We agree with this suggestion, provided that the habitual fusion of the notions of automaticity and unconsciousness is clearly debarred. Although Tzelgov himself is perfectly clear as to the claim, which we fully endorse (sect. 7.3), that evocative, automatic events are conscious, it should be remembered that both everyday language and influential psychological texts (e.g., Jacoby et al. 1997) use the terms *automatic* and *unconscious* interchangeably.

Maybe a more transparent terminology can be provided by the phenomenological tradition which, as **Bartolomeo & Dalla Barba** remind us, distinguishes between direct and reflexive forms of consciousness. For these commentators, the distinction is not aimed at excluding one of the two categories of events from the field of consciousness. Instead, they point out the interest of considering first-order, nonreflexive forms of consciousness in order to account for various phenomena, as we did ourselves. For example, Bartolomeo & Dalla Barba illustrate, using a situation borrowed from Merleau-Ponty, that there is no reason to attribute direct consciousness to the unconscious. On a related note, **Dulany** draws a distinction between first-order awareness and higher-order awareness, which "represents a first-order awareness of mental episodes." Schooler (2002) recently proposed a similar distinction, contrasting consciousness and metaconsciousness. This is not the place to elaborate on the kinships and differences between all of these proposals (Schooler [2002] provides a tentative clarification including the terms *metacognition, self-awareness, higher-order thought, metaconsciousness,* and other related concepts). For our purposes, it is essential to emphasize that our primary focus was undoubtedly on so-called evocative, automatic, first-order conscious events. The claim for self-organization is attuned to this level. The extent to which it is also relevant for reflective consciousness may depend on how reflective consciousness is conceived of. If one contends, as Searle (1992) has argued, that the mind is self-conscious roughly in the same way that it is conscious of the outside world, there is no reason not to apply the SOC model to metaconsciousness as well. However, we have not yet explored this issue.

Is our focus on first-order consciousness justified? A few commentators seem to think not. **Sun**, for example, claims that we have "greatly broadened the definition of the conscious." On a related note, **Dienes & Perner** argue for a more restrictive view. They introduce a distinction similar in many points to the preceding ones, although using another terminology. As detailed in Dienes and Perner (1999) in this journal, these authors consider fact-explicitness and attitude-explicitness as separate. Attitude-explicitness entails an additional level of explicitness, because the mental attitude one has toward the fact in question is explicitly represented. In line with a higher-order thought theory of consciousness, they posit that for a representation to be conscious, it must be both fact-explicit and attitude-explicit. As far as we understand their proposal, our unselective definition of consciousness entails that we consider that fact-explicit representations, to use Dienes & Perner's words, are conscious whether or not they are also attitude-explicit.

These specifications are certainly useful. Indeed, as **Carlson** has pointed out, "theorists who argue that what makes a representation conscious is having a higher order thought about it (e.g., Rosenthal 1993) might argue that the kind of minimal representation in awareness proposed by P&V is, in fact, unconscious representation." However, we see here nothing else than a terminological issue.[4] We borrowed what we judged to be the most common meaning of the word consciousness, namely phenomenal consciousness, but it is obvious that in so doing, we had no normative intent.

### *R2.2. Consciousness and control*

Although the notion of metaconsciousness is naturally related to the notions of monitoring and control, special comments are justified on the subject of control, because our article has been the source of misunderstandings on this matter. Indeed, assigning a key function to consciousness in adaptive behavior at the expense of the cognitive unconscious may be thought of as a plea for rehabilitating human control and freedom in the face of the current literature emphasizing the importance of unconscious influences.

As claimed by Reed and Johnson (1998), what they call



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

the "Consciousness as King" assumption entails "that we have conscious access and control over all important cognitive processes. We are, indeed, the captains of our ships." Along the same lines, **Destrebecqz & Cleeremans** coin the term "Commander Data" strategy to designate the strategy of those who deny any form of implicit cognition altogether. In keeping with this view, Destrebecqz & Cleeremans and **Jiménez** reason as though denying the existence of a cognitive unconscious would amount to positing that every aspect of mind functioning is under one's control, or at least that the mind is fully transparent to itself. Our own reference to Descartes (sect. 1.3.1) may be partly responsible for this idea, insofar as Descartes' claims that the mind is transparent to one's own consciousness (for a minor exception in Descartes' writings, see Guzeldere 1997, Note 34).

The preceding subsection should have made it clear that we identify consciousness neither with control nor with the transparency of mind. However, it is worth emphasizing that, in fact, our view leads to precisely the opposite position. This is because the way people consciously perceive and represent the world is determined by associative learning mechanisms to which no access is possible, whatever the introspective effort devoted to this endeavor, because they are not representational in nature. Let us take a very simple example in the conditioning area. After the paired presentation of two stimuli, S1 and S2, some aspects of S1 become associated to some aspects of S2. Let us consider that S1 is food and S2 an injection of a drug, and that a feature of the representation of S1, A (say its flavor), becomes associated with B, the illness elicited by S2. This is one of the most elementary illustrations of the self-organizing consciousness: The joint experience of two initially independent components leads to the modification of subsequent experiences because of the formation of a new unit. Flavor A now evokes an aversive reaction. In this sketch, everything mental is conscious. However, there is no possible control over the aversive feeling elicited by Flavor A, and the conscious experience of aversion has nothing to do with a capacity for knowing why this flavor seems unpleasant. The events responsible for this particular experience may possibly be traced by the subject if he or she has explicit memory of them, but the retrieval of the original experience is not causally responsible for the lived experience.

As rightly pointed out by **Redington**, this view, paradoxically, allows considerable scope for unconscious influences and conflicts, for example, with the position taken by Shanks and St. John (1994). The self-organization of conscious thought must not be understood as the organization of conscious thought under the control of the self. The principles governing self-organization are themselves unconscious.

### *R2.3. Assessing conscious contents*

Our focus on first-order, nonreflexive consciousness will help us to address a number of criticisms, as we shall see later. However, it also raises a thorny issue. **Kurthen** notes that conscious contents can be taken as a starting point for a scientific analysis of cognition thanks to their reportability, "by which they enter into meaningful behavior in scientifically utilizable settings." However, reportability is a definitory characteristic of second-order, reflexive consciousness, which involves, to quote Schooler (2002), "the explicit awareness of the contents of consciousness." Our focus on first-order conscious contents prevents us from using reportability as a general solution for the assessment of our object of investigation. Indeed, if there is some dissociation between having an experience and knowing that one is having this experience (Schooler 2002; see also Dienes & Perner 1999), or, maybe more commonly, some dissociation between having an experience and being able to report the nature of this experience verbally, what kind of data can validate our prediction about the content of conscious experiences? The question becomes even more problematic if one evokes a property of phenomenal experience that philosophers refer to as ineffability. This property means that the content of subjective experience cannot be taught to persons who have never had the same experience, at least to some degree. Thus, for example, it is known to be impossible to describe to a congenitally color-blind person what it is like to see colors (e.g., Jakab 2000).

Indeed, **Dandurand & Shultz** and **Manzotti & Sandini** wonder about the possibility of demonstrating that the representations generated in PARSER match the phenomenal experiences of subjects, as we claim. "Present approaches," write Manzotti & Sandini, "don't offer any ready-to-use methods, not even Francisco Varela's second-person neurophenomenology, nor Daniel Dennett's heterophenomenology (cf. Dennett 1991; Thompson 2001; Varela 2000). This problem is crucial when it's time to evaluate PARSER results." And they ask: "How do we know that PARSER is nearer to consciousness than other models?"

An immediate feeling is that the response is self-evident. It should be remembered that the only representations manipulated by PARSER, when exposed to an unknown language, are chunks of syllables, with each chunk containing a few syllables (such as *badu* or *dutaba*). Brent's (e.g., 1996) model, by contrast, posits that subjects perform all the possible partitionings of the corpus, then compute a complex mathematical maximization function to select the words from the other parts. We guess that anyone will acknowledge that PARSER provides a better approximation of people's conscious experience than Brent's model does. Of course, PARSER is unable to provide an exhaustive account of singular conscious experiences. The rich and multifaceted conscious content of a given individual at a given moment is certainly not predictable. But this situation, we believe, does not differ formally from that of other experimental sciences.

For example, the exact pathway of a dead leaf falling from a tree is also certainly not predictable. This does not mean that the phenomenon is of no concern to physicists. However, the goal of physics researchers is not to account for the fall of a single leaf but rather to investigate the physical principles governing this kind of event, such as the law of gravity, air resistance, the action of the wind, and so on. Likewise, we intend to make clear the principles underlying the progressive structuring of the content of conscious experience and we are tempted to say that some rough consensus about the structure of this experience can be sufficient at this level. The argument according to which such a rough consensus must be discarded because ultimately we cannot be sure about introspective reports, appears somewhat unfair.

However, we also feel that such a response is not fully satisfactory. Later we shall look at other cases, for example involving neuropsychological patients, where the nature of first-order conscious experience is actually not self-evident.


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Certainly new methods need to be developed in order to address in a principled way **Manzotti & Sandini**'s objection that conscious contents are not a suitable basis for a scientific investigation.

## R3. Processes and representations

Several commentators point out that our use of key words such as *process, representation,* and *computation* is unusual, sometimes inconsistent throughout the paper, and too fuzzy to make our position meaningful and falsifiable.

### *R3.1. Disentangling terminological issues from SOC's fundamental principles*

**Lambert, Ellis,** and **O'Brien & Opie** challenge the claim that a cognitive unconscious is an unnecessary postulate on the basis of examples drawn from the field of early vision. Lambert notes that:

> It is clear that even at the level of the retinal ganglion cells, the information encoded by the rod and cone receptors has undergone a variety of sophisticated transformations. For most cognitive scientists, these transformations are not simply regularities that may be modelled in computational terms, like the weather or planetary movements – they *are* computations, which extract information such as luminance and colour transitions from the retinal image.

Likewise, **Ellis** claims:

> Extensive processing occurs during the first 250 msec of processing, with or without the consciousness accompanying the 300P. These unconscious occipital transformations fit a computational paradigm: Cells in consecutive layers of sensory cortices analyze different features of perceived objects – lines, angles, shapes, colors, and so forth.

The fact that these transformations are unconscious is indisputable. To be consistent with our own viewpoint, we have then to posit that early visual processing involves neither representation nor computation. This position is seen as indefensible by some commentators.

First of all, it should be clear that we do not believe it to be incorrect to use the words "representation" and "computation" in connection with early vision. Terminology is mostly arbitrary, and we had no intention of making any normative statement in this regard. We even acknowledge that this particular terminology can be cogently motivated. Our claim is that one cannot justifiably conflate the representations and computation involved in early vision with the representations and computations that fill up our conscious experiences.

For example, the information conveyed by a cone in the retina about the external source of light that leads the cone to fire does not provide a representation in the same way that the conscious representation of an object does. Likewise, the transformations undergone at the level of the retinal bipolar cells, although they can be framed in terms of a summation between the central and the peripheral parts of their receptive fields, are unrelated, we argue, to the transformation we perform on represented digits when making mental calculations. In the target article, we used the term *processes* to designate the former transformations, and we reserved the term *computation* for the latter. To repeat, we believe this terminological choice to be arbitrary, but, because the distinction that we embody in it is at the core of our model, we considered it to be valuable to use different terms.

Are our terminological choices actually idiosyncratic and unjustifiable, as argued by **O'Brien & Opie** and **Lambert**? We believe not. Many other authors have noted that the recent extension of the use of the terms *representation* and *computation* to an increasingly large context makes these terms nearly meaningless. To quote Ramsey (1997):

> If all that is needed to qualify as a representation is for a structure to become active as the result of certain inputs, or go into states that are unique to certain inputs, then practically every neuron in our perceptual system is serving as a representation in this sense. . . . To recall an old but useful analogy, my digestive system goes into specific states (or combinations of states) that are to some degree unique to every meal that I eat. If someone wants to call these gastrointestinal states representations for this reason, so be it. However, it would be wrong to insist that interpreting the states in this way helps us to understand and explain digestion.[5]

### *R3.2. From processes/representations to the neural-mental distinction*

Beyond terminological matters, we come now to the fundamental issue, namely the need to make a distinction. The existence of a difference may seem obvious when we consider very contrasted events, such as the summation performed by retinal bipolar cells and the mental addition of two numbers. However, listing the formal criteria that underpin this distinction, in order to establish a clear frontier between what we call processes and representations, is far from easy. The difficulty increases when intermediary situations are considered. For example, what about the perception of the size of an object, a phenomenon that Rock (1984) described as an inferential process? (See above, sect. R1.1) An immediate response could be: Because calculating an object's size is not a conscious operation, it involves processes, not the manipulation of representations. However, as pointed out by **Carlson**, **Seger**, and **Dienes & Perner**, this line of reasoning is circular.[6] It amounts to defining the terms of representation and computation on the basis of their conscious status, thus making our thesis tautological. To avoid circularity, we need to delineate the terminological distinction that we make in a way that does not refer to conscious states.

Throughout the target article, we used the term *mental* to qualify representations and to distinguish them from processes. Does this avoid circularity? On the one hand, we can answer Yes, because this preserves the theoretical possibility of mental but unconscious events. However, on the other hand, it is virtually impossible to define mental events without making reference to our conscious experiences. In the most commonly accepted sense, mental activities are activities that look like conscious activities. Several commentators observe that this situation impairs the clarity of our model and complain about the fuzziness of the process/representation distinction. How can this problem be resolved?

In the target article, we often talk about a "powerful" or "sophisticated" cognitive unconscious to characterize the phenomenon we dismiss. In so doing, we may have given the impression that we view everything unconscious as underpowered and unsophisticated. This idea may have been strengthened by our repeated references to "associative processes," and even sometimes to "elementary associative processes." We will return shortly (sect. R3.4) to the confu-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

sion we have introduced here, but for our current concern, it is worth emphasizing that our position is different. The power and sophistication of the neural (unconscious) processes intervening in early sensory processing, including the retinal level, are unquestionable. The notion of "complexity" or related concepts is improper for distinguishing processes from representations.

**Lambert** rightly notes that the notion of "meaning" seemed to come close to our use of the term "representation." Thus, he proposes that our "central assertions could be recast as: (a) All semantic representations are conscious, and (b) there are no unconscious cognitive processes that operate upon semantic representations." Although this idea touches on an interesting consideration, we are afraid that the notion of "meaning" is itself too imprecise to be really helpful. Other commentators suggest an even more restrictive statement. According to **Sun**, we should refer to the idea of "symbolic" representation. Admittedly, limiting the notion of representation to symbolic contents would achieve our objective of excluding the transformations occurring in the early sensory systems. Indeed, in the classical sense, the notion of "symbol" conveys a notion of arbitrariness between the represented event and the symbol that stands for it, and it is certainly inappropriate to use the term for biological events. However, we feel that these solutions are unsatisfactory because they are excessively restrictive. Although it is certainly far from trivial to claim that there are no unconscious semantic or symbolic representations, our proposal is intended to be more general.

Another suggestion is made by **Dienes & Perner**, using the implicit/explicit distinction. Philosophers commonly define explicit representations as representations in which there is a term-to-term correspondence between the representing world and the represented world. Each feature of the representation stands for a particular part or property of the represented event. Nonexplicit representations may take various forms. Dennett (1982), for example, has distinguished between "implicit" and "potentially explicit" representations, which are actually nonexistent as such but can be logically inferred from explicitly represented information or retrieved by the system. He also refers to "tacit" knowledge, a form of knowledge that is embedded in the procedures of the computational device. Our proposal concerns only explicit representations, as **Dienes & Perner** rightly assume. This allows us, to quote Dienes & Perner, "to rule out the unwanted cases of activation patterns as cases of representation in a principled way." Clarifying this point would have also prevented some misunderstandings on the part of a few commentators. For example, **French** takes it for granted in the course of his argumentation that we agree with him that a concept can be represented in the brain "whether active or not." In fact, the concept of inactive representation has no place in the SOC framework. This restriction is inherent to the mentalist metatheory, in which, as pointed out by **Dulany**, a representation is not "stored" and then "retrieved." Any representation is built on the fly as a function of internal (the current state of the system) and external (the sensory input) contexts, and as a consequence its life span is limited to the current phenomenal experience.

Thus, stating that the representations we talk about, by contrast with processes, are meaningful and explicit may act as a useful specification. However, these features are insufficient by themselves, as shown, for example, by the existence of meaningful and explicit representations in symbolic computational models. Moreover, they ignore the point that we believe to be essential. In our view, the key property that makes it possible to differentiate between representations and processes remains whether these events are mental or not.[7] Using the terminology neural/mental instead of process/representation might have made our position clearer.

### *R3.3. The neural/mental distinction and the charge of dualism*

According a special status to mental events seems to amount to contrasting mental events with biological or, more precisely, neural events. This exposes us to the charge of dualism made by **Manzotti & Sandini**. As a consequence, we are also faced with the problems that can apply to any dualist position. **Velmans**, for example, denies that the mental can be causal, a claim that, if correct, is a strong argument against the SOC framework.

These questions are not specific to our proposal. They apply to any cognitive account, insofar as they involve causal relations between mental events irrespective of whether these events are construed as conscious or unconscious. Because this is not the place to deal with such broad issues, we limit ourselves to a few general remarks.

First, we do not believe that claiming that the mental has its own organizational principles commits us to a dualist position. We simply consider that there are multiple levels of explanation. What we mean here can be made clearer by enlarging the debate. Searle writes:

> 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)

The claim that the mental is endowed with specific organizational principles (e.g., relies on representations), which have to be distinguished from the organizational principles that apply at the neural level (the level of processes in our terms), is closely parallel to the claim that a goal in a football match is determined by specific causal events, which have to be distinguished from the causal relationships relevant in the movement of particles.[8] In both cases, the term *distinguish* does not mean that the two kinds of events are opposed or even independent, as substance dualists might argue concerning the mental/neural dichotomy. In the same way that football games involve the movement of physical particles, mental events involve neural activities. A conscious visual experience obviously involves the processes operating at the level of retinal bipolar cells. As we claimed in our article (sect. 1.3.1), processes and representations are "linked like the head and the tail of a coin." However, biological laws are no more relevant for the description of mental events than the principles governing particle movements are relevant when describing football games.

Of course, Searle's analogies serve more to illustrate the generality of the problem raised by **Manzotti & Sandini**, **Velmans, Dienes & Perner**, and others than to solve it. Returning more precisely to the causal power of con-


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

sciousness, we referred in the target article (sect. 9.2) to the neural interactionism proposed by Roger Sperry as a position compatible with our framework. We mentioned that the principle of downward causation inherent to this position now receives additional support from the field of dynamic system theory (e.g., Thompson & Varela 2001). Because this position is certainly far from consensual, it is worth emphasizing that some apparent counterarguments may not be as firmly grounded as might have once been thought. This is the case in particular of the statement that the brain starts to prepare for motor actions about 500 msec before our conscious intention to move. This statement derives from the famous results obtained by Libet (e.g., Libet et al. 1983), which have been accepted somewhat uncritically by so many researchers (e.g., Dennett 1991). However, these results have been the subject of a number of reappraisals, most of them conducted during the last few years (see, in particular, the special issue of *Consciousness and Cognition* 11 [2002]). Although it is fair to mention that Libet still maintains his original position against his numerous opponents, we also consider that Libet's findings and interpretations are as yet insufficiently grounded for it to be possible to draw firm conclusions as to whether consciousness is devoid of any causal power. However, we are ready to acknowledge that this kind of issue is far from settled and that a major collective effort involving all the fields in the cognitive sciences is required in the future.

#### *R3.4. Neural processes versus associative processes*

A recurrent source of confusion in the target article stems from the fact that we used the word "processes" to designate both the "neural processes" we referred to in the above subsection and the "associative processes" that are central to our position. We suspect that several criticisms made by **O'Brien & Opie** and **Rivera-Gaxiola & Silva-Pereyra** are due to this confusion. Indeed, this confusion obscures (1) the fact that the whole range of biological processes involved in the brain cannot be reduced to elementary associative mechanisms; and (2) more important for our purpose, the fact that associative processes operate on conscious mental contents, which is a central tenet of our framework. The self-organization of the mental relies on the existence of associative processes bearing on mental contents and not on neural informational states, which are themselves relevant at another level of organization.

### R4. Do we need an "intermediary level"?

In keeping with William James's and Dulany's mentalism and philosophers such as Merleau-Ponty and Searle, we reject a "middle-level" concept falling between the level of biology and the conscious level. By contrast, several commentators emphasize the need for an intermediary level of explanation. Although this is undoubtedly an oversimplification of their standpoints, let us summarize their proposals as a three-level causal chain, reminiscent of a Fodorian architecture, such as this:

> Biological level → Unconscious mental level → Conscious mental level

The need for the intermediary level may be set in the context of phylogenetic evolution or considered as relevant for the organization of the human mind.

#### *R4.1. The evolutionary argument*

**French** starts his commentary with "the lovely aphorism on which Darwin based his work on evolution," namely: "Natura non facit saltum" (Nature does not make leaps). However, he seems to use this quotation in a metaphoric way, because in fact his commentary concerns the organization of the mind. In contrast, **Litman & Reber** genuinely address the phylogenetic evolutionary level. They contend that:

> [our] point of view assumes that all observable rule-governed behavior and all abstract memorial representations are the products of a recently evolved consciousness . . . and that no proto-rule functions or proto-abstract representations would be expected in other species or in our ancestors. Contemporary models of evolutionary biology simply do not allow for these kinds of rapid evolutionary saltations, certainly not within the time frame dictated.

In fact, the claim that consciousness is a "recently evolved" function which other species or our ancestors lack or lacked comes from **Litman & Reber** themselves. Starting from this postulate, Litman & Reber infer that some unconscious "proto-abstract" representations must have existed in our ancestors in order to comply with the principle that evolution is gradual. But the need for an unconscious, intermediary level in the evolutionary story is nothing other than the other head of their initial postulate, namely, that consciousness is a late arrival. Now, inasmuch as "first-order," nonreflexive forms of consciousness are concerned, this standpoint is in no way compelling. Our repeated references to the literature on learning in animals (e.g., sect. 2.2.2, Note 11), as well as the close link we introduce between the concepts of attention and consciousness (sect. 2.2.1) should have made it clear that we do not subscribe to a view that, in our opinion, confounds consciousness and language. Given that our own view on this matter espouses the position that is prevalent in the contemporary scientific community, we did not think it necessary to expand on this issue. Nevertheless, if one does not accept the idiosyncratic standpoint that consciousness is specific to humans, Litman & Reber's objection no longer holds. Elementary forms of consciousness may have appeared early in animal evolution and then have evolved gradually.

As an aside, we find the charge of **Litman & Reber** that we ignore the principles of evolutionary biology highly paradoxical. The general stance of our paper, such as the weight given to associative learning principles, and moreover its focus on self-organization, make the SOC model much more compatible with evolutionary principles than the classical computational view (see, especially, sect. 3.3). We also noted (see Note 9 of the target article) how evolutionary principles could also have shaped the properties of language. **Wang & Ke** cogently discuss this aspect and provide exciting suggestions to complete and extend the SOC model in line with recent developments regarding the phylogenetic and ontogenetic development of language.

#### *R4.2. The structure of the mind*

Let us turn now to the need for an intermediary level in the formation of conscious representations from the neural apparatus. **French**, **Jiménez**, **Sun, Ellis**, and **Destrebecqz & Cleeremans** affirm the compelling nature of such a statement. For example, Sun opposes our SOC model with his BUC model, where BUC stands for "bottom-up con-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

sciousness." In his view, consciousness necessarily emerges from somewhere in a bottom-up fashion. This "somewhere" is the level of unconscious representations, the existence of which he construes as self-evident. Likewise, Ellis advocates the necessity of "potential representations – activities that will lend themselves to use by the organism in representational action if the occasion should arise." In a similar vein, Destrebecqz & Cleeremans claim that "attention selects the strongest representations that are active at some point in time to form the focus of awareness." This kind of claim supposes a selective device, which is the "organism" for Ellis and "attention" for Destrebecqz & Cleeremans. The purpose of this device is to sort out what is useful for the task at hand among all the possibilities that our unconscious apparatus has prepared for us in "the deep waters of our brains" (Jiménez).

In the same way that we address the evolutionary argument above, we argue in this case that the very existence of an intermediary level, far from constituting theory-free evidence, is in fact nothing more than a logical consequence of initial postulates about the structure of the mind. If consciousness is viewed as a device that gives access to some internal states of the system, as in **Sun**, **Ellis**, and **Destrebecqz & Cleeremans**, and more generally in the computational framework (see sect. 1.1), the need for a cognitive unconscious follows. Without a cognitive unconscious, there could be no response to the question: Access to what? However, if one starts from the opposite postulate, the same approach becomes meaningless: Far from fragmenting a problem into individual steps, postulating an intermediary level amounts to introducing an additional problem that exists only in researchers' minds. In some sense, the concept of SOC is entirely intended to show why this intermediary level, from which consciousness is assumed to draw its content, is a useless postulate. By transposing the functional dynamic of the system from the unconscious to the conscious level, the SOC deprives the cognitive unconscious of any functional justification. Note that if one accepts our postulate that consciousness is an intrinsic property of the mental, introducing the intermediary level appears not only meaningless but even bordering on the absurd. Indeed, it is tantamount to believing that in order to study physical objects, one property of which is the possession of mass, the best initial approach is to look at objects that have no mass!

In the discussion above, we assume that the point of debate concerns the internal organization of the mental. In so doing, we assume that this organization is without consequences for the nature of the first arrow in the three-level chain above. This arrow, which separates the neural and the mental level, represents what is often referred to as the "explanatory gap." As we see it, we are faced with the explanatory gap in the same way as authors who propose an intermediary level. However, some aspects of **French**'s commentary could suggest that the transition from the neural level to the mental level is no longer a problem when the mental unconscious is concerned. In our reading of **O'Brien & Opie**, a related idea is present. The operation of unconscious, structure-sensitive rules is thought of as necessary to address the problem of the physical implementation of cognitive processes. We see nothing compelling in these claims. In our view, the use of expressions such as "neural representations," as in French's commentary, does not in any way bridge the gap existing between neural connectivity, whether in brain or in artificial networks, and mental representations. Likewise, we fail to see how invoking the rules and symbols of classical computational theory of mind could facilitate a biological implementation of mental events (the opposite conclusion would seem to be more justifiable).

### *R4.3. The notion of "graded representation"*

In the three-level causal chain above, the transition from the unconscious mental level to the conscious mental level is represented as a one-step process. This does not do justice to the proposals of several commentators, according to which this transition is graded. It should be clear that the debate between a one-step and a graded transition is irrelevant in the context of our paper: Because we obviate the need for an intermediary level altogether, the problem of the form of the unconscious/conscious transition does not arise. However, it appears that some commentators consider (often tacitly) the gradedness of the transition to be proof of the existence of a cognitive unconscious. We believe this argument amounts to making an incorrect inference from a correct observation.

The correct observation concerns the fact that mental representations may differ in their strength or quality. Much of **Destrebecqz & Cleeremans**'s commentary constitutes a plea for the existence of graded representations. We fully agree with them on this point. In fact, as rightly pointed out by **French**, our paper was quite explicit on the fact that we accept the notion of graded and partial conscious representations. This is a trivial consequence of an association-based view of representation: Associations are strengthened through repetition and vanish with decay and interference. In keeping with these general principles, the notion of graded representation is a fundamental feature of PARSER, in which representations are ascribed a weight.

However, the key question is: What is the status of weak representations along the conscious/unconscious dimension? For example, are weak conscious representations close to the unconscious level? **Destrebecqz & Cleeremans** are persistently ambiguous in their commentary. They seem to acknowledge that weak representations are a part of conscious experience, for example, when they suggest that the "clear first-person evidence that our phenomenal experience . . . includes a fringe of attenuated contents" is a proof of the notion of weak representation. However, elsewhere they mention "the weak representations characteristic of implicit cognition," thus assimilating weak conscious representations with unconscious representations. This assimilation is latent in other commentaries as well. For example, **Houdé** speaks about "a hierarchy of levels of consciousness (*and thus* of partial unconsciousness)" (our emphasis). Likewise, **Wolters & Phaf** speak at times of a "relative level of consciousness" or "less conscious" events in order to designate what they construe as unconscious stimuli. This kind of conception translates what is a common intuition. For example, when we experience difficulties in retrieving memories or when a word is on the tips of our tongues, we spontaneously interpret those experiences as if the memories or words do exist somewhere in our brain as full-blown representations, but in locations that we have difficulty in finding.

We argue again that this view is a presumably unintentional commitment to the classical framework, in which


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

conscious representations emerge from a richer cognitive unconscious. This leads us to believe that what is unclear in consciousness necessarily exists in the unconscious. Because consciousness is construed as a device providing access to some mental contents, the lack of clarity is attributed to a failing in the access process. However, it is worth emphasizing that there is no additional support for the notion of unconscious representations in these claims: They amount to nothing more than the initial postulate in a new guise. The notion of low-quality conscious representations is perfectly compatible with the denial of a cognitive unconscious.[9] For example, the experience of laborious retrieval of memories, or tip-of-the-tongue phenomena, can be conceived of as a difficulty in constructing new representations, given the current state of the system, instead of as a difficulty in retrieving a ready-to-use representation hidden somewhere in our heads.

We even believe that instead of providing proof of the existence of a cognitive unconscious, the graded nature of consciousness actually provides an argument prompting the reinterpretation of certain data that seem to support a conscious/unconscious dissociation. For example, Munakata (2001) shows how the notion of graded representation may act as a replacement for the idea of system dissociation: Some tasks may require strong representations, whereas other tasks may be successfully performed on the basis of weak representations. Again, it is essential to distinguish between the low-to-high-quality dimension and a putative conscious/unconscious dimension, and not to consider evidence for the former as support for the latter. Having said this, the notion of graded consciousness is also limited and potentially misleading. Indeed, it may suggest that conscious experiences vary only along the single dimension of clarity. The idea that conscious experiences can differ in nature and contents, as emphasized in section R.2, is certainly a more powerful explanatory concept, as we shall see later.

### *R4.4. The intermediary level as a connectionist network*

Connectionist models account for rule-following behavior without implementing any rules in the system. This is a major point of convergence with the SOC framework and one that would be sufficient to place our proposal within the connectionism camp if the current theoretical landscape turned out to be reduced to a dichotomy between formal and connectionist models. The points of agreement go even further. Connectionist models elegantly implement the dynamic principle that the system changes as a consequence of its own functioning. **Redington** rightly also notes that representations, conceived of as the pattern of activation over units, occur only in a transient way in response to the incoming information. Indeed, connectionist networks avoid postulating the storage of a huge amount of symbolic, ready-to-use information which is typical of the classical framework, as the SOC model does. On the basis of these convergences, a few commentators (**O'Brien & Opie**, **Destrebecqz & Cleeremans**) who express sympathy with several of our arguments point out that connectionist models implement some principles of the SOC framework but possess additional advantages. This subsection is aimed at assessing why connectionism, despite the above-mentioned similarities with the SOC, implements a radically different conception of the mind.

Any discussion of connectionism is made difficult by the fact that the term does not convey a unified corpus of methods. Suffice it to evoke the differences between distributed and localist modes of representation or alternatively between supervised and unsupervised algorithms, the choice between which is obviously crucial when discussing the relevance of connectionist models to SOC's principles. Even within a given configuration, researchers disagree about the way the components of their models should be interpreted. Regarding, for example, the notion of representation, some authors (among the commentators: **French**; **Dienes & Perner**, and **O'Brien & Opie**) give the status of representations to both the pattern of connection weight between units and the activation pattern (although often distinguishing the two forms as respectively implicit and explicit), others (**Destrebecqz & Cleeremans**, **Redington**) consider only the activation pattern as relevant, whereas still others argue that talking about representation in connectionist networks is unwarranted (e.g., Ramsey 1997). A property such as distribution is seen either as an obstacle to invoke the existence of genuine representations (Ramsey 1997) or as a property that makes it possible to mimic conscious experience (Lloyd 1996). Given the limited space available for this response, this discussion is limited to a few general aspects.

In the target article (Appendix A), we mentioned two features of connectionist models that depart from the SOC framework. The first is that the to-be-learned material is coded via the same input units throughout training sessions, whereas one of the main tenets of the SOC is that perceptual primitives continuously change with experiences. It could be argued that a hierarchy of networks, with a second network taking the representations shaped by a first network as its input, could address the problem. However, we pointed out that representations are not formatted to serve as new coding primitives. Second, partly as a consequence of the first characteristic, it is difficult to implement the idea that associations apply to increasingly complex representations.

In their commentary, **Dandurand & Shultz** mention that one model they have developed, the Knowledge Based Cascade Correlation model (KBCC), addresses these two problems. The basic idea of their model is to expand the network topology, when this is required for the processing of a complex task, by recruiting subnetworks that have previously been trained on simpler tasks. After this expansion, the connection weights of the subnetwork are not affected by further learning, thus leaving intact the knowledge of the simple task. Dandurand & Shultz cogently outline two points that remain as potential difficulties if one hopes to implement PARSER in a connectionist architecture. In their current instantiations, KBCC networks follow a supervised algorithm and, more problematically, they need external interventions in order to plan the successive steps of learning. We also foresee another difficulty. A given network can indeed profit from the representations built by one or more subnetworks, but the representations in question are defined by the connection weights between units and not as the transient activation patterns generated by a given input. They are the "implicit" representations in **Dienes & Perner**'s and **O'Brien & Opie**'s terminology. It is not clear whether this architecture makes it possible to implement the dynamic, online integration of explicit representations that is fundamental to SOC's principles.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Irrespective of the developments of the KBCC models, however, there is an irreducible difficulty facing the implementation of the SOC principles within a connectionist architecture, whatever it may be. This is simply that connectionist networks represent, in fact, a certain way of conceiving of the intermediary level, which we intend to remove from the architecture of mind. Whereas the SOC model (1) accounts for conscious contents; and (2) gives them a key function in the learning process, connectionist models fulfill neither of these two functions.

Regarding the first point, let us consider, as an illustration, the word extraction issue developed at length in the target article (sect. 3). It can be argued that after learning has been completed, a simple recurrent network accounts for the conscious perception of words (assuming additional computations devised to infer the words from the sawtooth-shaped distribution of errors). However, this is no longer true if one considers the conscious percepts during learning. Conscious experiences do not evaporate while learning progresses: Subjects do have discrete percepts, which cannot be accounted for by the nearly random distribution of the weights of the output units. This inability was the object of our commentary (Vinter & Perruchet 1999b) on the *BBS* target article of O'Brien and Opie (1999a).

The second point is even more important. Conscious percepts have no place in the learning process of a network. As pointed out by **Dandurand & Shultz**, "attention is not a cause of the model's successful learning but a result of learning – the networks learn what features of the stimuli to attend to." Thus, conscious percepts are both unexplained and useless, that is to say, they are excluded from any causal chain. Connectionist models are necessarily committed to a view in which consciousness provides a somewhat epiphenomenal access to the end product of learning. The dynamic of the system is due to unconscious computations, the operation of which is completely disconnected from phenomenal experience. The computations involved in connectionist networks are certainly different from the symbol manipulation involved in the classical framework, but they are still computations. Talking about connectionism, **O'Brien & Opie** claim: "There is a path that runs midway between classical computationalism and mentalistic associationism; a path that avoids the classical unconscious *without* abandoning computationalism." Although understanding the attractiveness of this path, we do not wish to take it, because we believe that there are good reasons to throw out not only the classical bathwater but the computational baby as well (to borrow O'Brien & Opie's title).

This being said, we believe that the success of connectionist approaches in a growing number of domains is a useful step because it shows that the distributional properties of the material (frequency, co-occurrence, contingency, and other statistical descriptors) provide sufficient information to achieve the task. In so doing, connectionist approaches demonstrate that the formal computations envisaged in the classical information-processing framework are not necessary. However, they do not demonstrate that human subjects exploit the distributional properties of the material through the analytic procedures embedded in the connectionist algorithms. Our guess is that where connectionist models are successful, PARSER-like models should also succeed, thanks to the exploitation of the distributional properties inherent in the self-organization of conscious thought, without involving any intermediary level of causality.

### *R4.5. Does PARSER actually imply "intermediary levels"?*

One specific argument put forth by **Destrebecqz & Cleeremans** and **Jiménez** in support of the necessity for an intermediary level of representation is based on the claim that Parser would use this form of representation, in contradiction to its own underlying principles. For example, Destrebecqz & Cleeremans claim that:

> It is crucial to realize that sub threshold units in PARSER, far from being impotent, are in fact central to its functioning. To see this, one needs only to consider that a unit that is just below threshold has a greater probability to shape processing on the next episode than low-strength units. Subthreshold cognition is thus functional after all, if only potentially so.

Although we are afraid that dealing with this argument demands some technical considerations about PARSER, a clarification is worthwhile.

**Destrebecqz & Cleeremans** are right in claiming that subthreshold units are essential to PARSER's functioning. Subthreshold units consist of a few syllables the concatenation of which is fragile. This is because they have not been perceived together frequently enough to compensate for the effects of decay and interference. They are unable to shape immediate perception. For example, if the unit *dutaba* is in such a provisional state, subjects may still perceive *du* and *taba* separately instead of considering *dutaba* as a perceptual primitive. However, if subjects perceive *dutaba* as a whole, thanks to the fortunate association of the two primitives *du* and *taba,* the resulting weight of *dutaba* will be stronger than if this unit was not previously in a subthreshold state.

**Destrebecqz & Cleeremans** wrongly consider the threshold in question to be a threshold of consciousness, with suprathreshold units being conscious and subthreshold units being unconscious. In fact, none of the units stored in the percept-shaper are genuine representations, and hence the question of whether they are conscious is irrelevant. Their status is only dispositional. The use of a storage device is a computational ploy aimed at simulating the progressive changes in the conscious perception of a given input throughout learning (see Appendix A). However, another objection arises. Indeed, because subthreshold units are unable to shape the content of the percept when faced with the raw material, they would represent changes in processes that have no effect on the phenomenal experiences. This possibility appears to be inconsistent with our claim that processes and representation are like the head and the tail of a coin. Along the same lines, **French** challenges us on the basis of a thought experiment in which certain homeopathic modifications in the processes involved in learning a concept improve subsequent learning of the same concept later in time without altering the conscious experiences of the learner.

We acknowledge that our position with regard to this matter is not yet clear. One way to deal with the objection would be to claim that a change in processes always has a counterpart in the subject's conscious experience, however subtle the modification in conscious experience is. For example, although the informational content remains unaf-


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

fected, the perception of chunks matching the subthreshold units in PARSER could be accompanied by the same feeling that we have when we recognize an event that we have failed to recall. In a still finer way, perception may be accompanied by a vague feeling of déjà-vu or may simply be more fluent. These changes in conscious experiences are irrelevant for PARSER's functioning but could be relevant in other contexts. It is known, for example, that processing fluency can affect various judgments (e.g., Jacoby & Kelley 1987). If one endorses this view, our response answers the argument put forward in the preceding subsections, namely, that conscious experiences of various strength or nature are erroneously conceived of as instances of a dissociation between conscious and unconscious states. However, admittedly, this hypothesis may turn out to be unfalsifiable, as **French** points out.

## R5. Language

Up to now, the issues we have considered have related to the general lineup of the SOC model. In the remainder of this response, we address the commentaries regarding specific domains of application. In keeping with the structure of the target article, we will deal successively with language, transfer and reasoning, creativity, and subliminal phenomena.

### *R5.1. Learning the meanings of words*

**Yamauchi** estimates that during the 18-months-to-6-years age period, children learn about 10 new words a day. Associative learning theory, he argues is unable to account for the fact that children come to learn so many words. He reasons that a scenario in which associations are formed as a by-product of the attentional processing of the word-world pairs cannot work, because repetition of an event tends to distract the attention of the learners. In a nutshell, the formation of associations needs both attention and repetitions, but repetitions reduce attention, thus making association impossible.

If **Yamauchi**'s conclusion is correct, this would mean that associative learning does not exist at all, a proposal, we guess, that nobody would be ready to endorse. Where is the trap? The response is twofold. First, it is wrong to postulate that any association requires a huge amount of repetition. In the area of animal conditioning, one-trial learning is a well-documented phenomenon, even if the reasons that make some associations (e.g., smell/illness, tone/electric shock) far easier than others (smell/electric shock; tone/illness) remain obscure. In word learning, many studies (e.g., Carey & Bartlett 1978) have also reported very rapid learning, sometimes called fast-mapping, although, here again, the question of what events can and cannot be fast-mapped is still open. But let us take the case where learning requires many repetitions. Yamauchi's reasoning does not hold for a second reason, namely, the fact that to a large extent, habituation is itself a consequence of the formation of association instead of being an autonomous process preventing the formation of associations. Indeed, attention does not fade away for an isolated event as such but instead because a target event is associated with other events, which make it predictable. To illustrate, even a very common object, say a fork, draws attention to itself if it is in an unusual context, for example in the place of a necktie. This explains why associative learning is a ubiquitous phenomenon. By definition, irrespective of whether S1 and S2 are familiar events in themselves, the pairing of S1 and S2 is new for the subject and hence goes on to capture attention up to the point at which an association is formed. Habituation will occur for the S1–S2 pair only when an association between S1 and S2 is learned. As a consequence, the SOC model does not need the intervening subsystems that Yamauchi claims are necessary to maintain attention despite repetition: Consciousness is definitely a self-organized system.

This being said, we do not conceive of vocabulary learning on the model of a giant conditioning or paired associates experimental session, consisting of the sequential acquisition of piecemeal associations between pairs of well-identified items. We agree that the issue is much more complex. However, the question is again: Can additional complexity be encompassed within a full-blown association-based model, or is it necessary to introduce some sophisticated cognitive unconscious, such as unconscious inference or hypothesis testing operations guided by inductive constraints (e.g., Markman 1990)? Again, our guess is that the former explanation is sufficient.

As an illustration, let us consider the fact that the rate of learning differs across lexical classes in a way that mismatches the distributional properties of the verbal input. Notably, children learn nouns faster than they learn verbs, although adult speech contains both nouns and verbs. This fact seems at odds with an association-based view. However, Gillette et al. (1999) propose and validate an explanation that obviates the recourse to extraneous principles. Their proposal is that the information entering into associations becomes available to children serially and depends on the formation of prior linguistic representations. At least some nouns can be identified on the basis of their extralinguistic context of use, hence allowing word-to-world pairing, whereas the identification of verbs would also require an examination of the linguistic contexts of use, hence involving sentence-to-world pairing. Such a scenario obviously requires significant refinements with regard to the simplistic mechanisms implemented in PARSER, but we believe that it provides a nice illustration of the self-organization of conscious thought.

### *R5.2. Syntax and semantics*

Several commentators state that the SOC model is incomplete in its present instantiation. **Parisse & Cohen**, for example, note that we consider only oral language. Because oral utterances are always transient, they are ill-suited to revealing the efficiency of certain of the principles on which the SOC relies, such as the possibility of using sensory data as an "'outside memory." Applying the SOC principles to written language would provide a better basis for an account of how learners become able to form representations that are isomorphic to the structure of their language. Moreover, it is possible, according to Parisse & Cohen, that access to written language results in a change in the level of analysis. From their initial, epilinguistic knowledge of language, children might exploit the presence of written material to gain metalinguistic knowledge, that is to say, knowledge about the nature and the structure of their own



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

language. This suggestion constitutes a promising development that Parisse & Cohen construe as fully compatible with the SOC framework.

The question now is: Can such amendment and others ultimately account for the whole area of language within an enlarged version of the SOC framework, or are there principled limitations in the SOC framework that a priori doom any such attempt to failure? Although **Wang & Ke** judge that "the SOC model is a plausible framework in addressing the ontogenetic development of language in individuals," other commentators are more skeptical. The issue becomes especially thorny when complex linguistic skills are considered. Syntactic knowledge provides a relevant example. **Rivera-Gaxiola & Silva-Pereyra** observe that even 3-year-old children are sensitive to syntactic violations in sentences and even in cases where the semantic content cannot help. In this regard, **Redington** asserts that "symbolic accounts of grammar possess far more explanatory power than any associative challenger." We agree with this claim but we add two remarks.

First, the key point is not the relative power but the question of whether the challengers are sufficient. Associative approaches to language are often believed to be bounded by principled arguments, which should have been the object of quasi-mathematical demonstrations. Thus **Rivera-Gaxiola & Silva-Pereyra** evoke what they call "the old-age argument of the 'poverty of stimulus'." The shortcomings of this argument are as follows. Assessing a priori whether the input provides sufficient information is necessarily dependent on one's conception of mind because it is necessary to specify what the information must be able to account for. Suppose, for example, that the information displayed in the input is insufficient to permit rule abstraction. This limit is relevant if rule knowledge is construed as necessary for language comprehension and production, as is the case in the classical view. However, the same limit may be irrelevant in an alternative framework. To illustrate, a special instantiation of the poverty-of-stimulus argument is the lack of negative evidence in normal speech input. Of course, negative evidence is needed for hypothesis testing and the inferential operations postulated in the classical approach. On the other hand, negative evidence is nearly irrelevant, if not detrimental, to learning in an associative framework. (Note that the reverse is also true: Information can be well suited for rule inference but unsuitable for association-based approaches; for example, if frequency information is lacking). The circularity is patent: the poverty-of-stimulus argument is used to support a position, whereas the validity of this argument is itself dependent on one's commitment to this position. To recapitulate, we believe that the "poverty-of-stimulus" argument is meaningless unless complemented by a theoretically rooted response to the question: Poor for what?

The SOC framework amounts to postulating that the distributional information provided by the input is rich enough to make us sensitive to the syntactic structure of the language, whatever its value as a database for a grammar-abstractor. Is this hypothesis plausible? In the target article (sect. 5.3), we alluded to the fact that some current streams of research in psycholinguistics circumvent the need for syntactic rules in language understanding and production. Although firm conclusions are admittedly premature, we shall see later (sect. R8.2) why we think it heuristic to make a bet in this regard.

Second, we believe that **Redington**'s claim about the power of symbolic accounts of grammar might turn out to be an overstatement if generalized to any language skills. Indeed, despite the high cost of a symbolic unconscious, some human abilities seem to lie beyond the scope of a computational approach. A nice example is provided by **Velmans**. Velmans presents a task borrowed from Green (1981) which consists of sorting 12 sentences differing in vocabulary and surface forms into pairs conveying similar meaning. Humans perform this task with little difficulty. Although Velmans cites this example as an argument against our own framework, he rightly notes that no program would perform above chance level in this task. This example helps us to restate the point at issue. The classical computational framework is certainly well suited to account for many linguistic competences, and it is no surprise that researchers have focused on the study of these competences. Syntactic knowledge, which has received so much attention from the early 1960s onwards, illustrates the point.[10] But the classical framework may be under-equipped to deal with other aspects. Although it is hardly more than a speculation, it is not fanciful to think that the SOC framework, thanks to its focus on the building of a transient but complex and multifaceted conscious representation, may provide a better metatheoretical approach for thinking about the kind of situation investigated by Green.

Another aspect that could be better accounted for by our approach, as pointed out by **Wang & Ke**, concerns individual differences. Wang & Ke remind us that individuals differ greatly in their preferred constructions. Children often follow an idiosyncratic path in acquiring their language. This variability does not fit well with a conception according to which language acquisition adheres to some preestablished building plan. By contrast, the idea that acquisition can follow different pathways as a function of the way the different components interact is a basic characteristic of self-organized systems (Thelen & Smith 1994). To summarize, we believe that (1) the alleged impotency of association-based or distributional approaches is true only if one endorses a ruled-based view of the mind, hence making the argument circular; and (2) the alleged superiority of the classical approach could be limited to the phenomena on which psycholinguists have focused since the emergence of the computational framework.

## R6. Accounting for transfer and reasoning

### *R6.1. Transfer and transitivity*

Both **Phillips** and **Tzelgov** describe specific experimental results that they consider to be difficult to encompass within the SOC model. Let us first consider the challenge offered by Phillips. Subjects are first presented with pairs of stimuli which are displayed either on identical or on different colored backgrounds. The subjects are then presented with one uncolored probe pair, and they have to select from among other uncolored pairs the one that has the same symbolic relation as the probe pair. According to Phillips,

> transfer is achieved with the capacity to bind arbitrary pairs of stimuli to symbols (representing the symbolic relations SAME and DIFFERENT); and depending on the decision process, either the capacity to match retrieved pairs from relational symbols or match retrieved relational symbols from pairs.



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

The perception of a stimulus and the color of its background within a unitary conscious experience makes their association possible – and even mandatory – according to the fundamental tenets of associative learning. This means that when subjects are later exposed, during the transfer test, to uncolored stimuli, they are able to remember the background colors seen during the study phase. Because their task consists of selecting a pair on the basis of symbolic relations, they select the pairs on the basis of whether they were displayed on the same or different backgrounds during the study phase. In fact, we do not see where the challenge involved in this task lies. If it is that colored backgrounds are not displayed again during the test phase, it is worth recalling that the momentary conscious representation is not limited to the current sensory input (the opposite claim would amount, for example, to denying the existence of explicit memory). If it is that subjects are able to abstract relations such as sameness from a pair of displayed or remembered stimuli when they are asked to do so, it should be recalled that we have explicitly discussed this ability in our target article (sect. 6; see also sect. R6.3).

The situation described by **Tzelgov** is more problematic for our model. Subjects are first shown pairs of arbitrary figures, each figure corresponding to a number from 1 to 9. Through the feedback they are given, they learned to perform magnitude comparison. Importantly, only contiguous "artificial numbers" are presented for training. The figures are then displayed in various sizes, and the subjects are asked to perform a magnitude judgment on their physical size. The point is that a size congruity effect is obtained on nonadjacent pairs. "Explaining these findings," Tzelgov argues,

> requires assuming that the training resulted in generating a representation that not only aligns all the arbitrary numbers on a single number line but also includes as a default (or automatically results in) the activation of their relative magnitudes whenever a pair of such numbers is presented.

If one takes for granted that artificial numbers have been arranged along a continuous line during the study phase, the automatic activation of their relative magnitude does not seem specially puzzling. In fact the problem becomes similar to the case of Arabic numbers. Again, the fact that conscious representations are not limited to the current sensory input but include components that have been associated with these stimuli in previous conscious experiences, irrespective of whether they are relevant for the task at hand, seems to provide a satisfactory account. However, how is it possible to account for the fact that the simple presentation of adjacent numbers permits the formation of a representation linking all the numbers to each other?

One possible scenario is the following. When subjects have to learn, say, that F3 is smaller than F4 (F3 indicates the figure that is arbitrarily linked to the number 3, and so on) just after having learned that F3 is larger than F2, it is quite possible that subjects use a strategy which consists of forming a conscious representation of the linear order F2 $<$ F3 $<$ F4. Of course, this strategy is not dictated by the instructions. But subjects may use it as a ploy to rehearse mentally in a holistic way the piecemeal information provided by isolated trials. Generalizing this strategy to a longer list of items could provide a unified representation of the series of figures. Accordingly, the question of whether this kind of interpretation works well or not warrants further investigation. However, it would appear to be necessary to test this or other alternative interpretations before accepting **Tzelgov**'s inference that "SOC has to include mechanisms that mimic intentionally applied algorithms."

### *R6.2. Reasoning*

**Houdé** argues that an experimental situation on its own shows how initial unconscious reasoning errors can be inhibited through specific training, allowing subjects to follow the correct deductive rule relevant for the task at hand. The notion of "unconscious reasoning error" is obviously at odds with the SOC framework, just as the notion of unconscious correct reasoning would be. Both presuppose that some mental computation mimicking conscious reasoning, irrespective of its correctness, has occurred.

Let us examine the situation used in Houdé et al. (2000), which allegedly reveals an unconscious reasoning error. Subjects have to falsify a rule such as "If there is not a red square on the left, then there is a yellow circle on the right." Most subjects respond "red square on the left, yellow circle on the right." This response is a nonconscious error in the sense that subjects are unaware that their response does not conform to the rules of formal logic. Obviously, we have no problem with such a reading. However, it is not the reading proposed by **Houdé**, who, in keeping with Evans's theory and more generally with the mental logic framework, considers that the response testifies to erroneous unconscious reasoning because of a perceptual bias. However, other interpretations are possible within the alternative framework based on the online construction of mental models, which we argue in favor of in our target article (sect. 7.1.1). For example, subjects may simply build two conscious perceptual models with a yellow circle on the right and a different event on the left (e.g., nothing vs. a red square), then consciously observe that one is the reverse of the other.

After this phase, subjects receive special training in a similar task (the Wason selection task). They are told (1) to avoid falling into the trap of considering only the events mentioned in the rule; and (2) what strategy they have to follow, which essentially consists in examining all the possibilities and eliminating the wrong ones. Then they perform the initial task once more, and most subjects now provide the correct response. In the example above, a correct response, for instance, is: "blue square – green diamond." **Houdé** argues that this performance is made possible thanks to the fact that subjects are able to inhibit the erroneous initial response. The idea of inhibition is indeed consistent with the hypothesis that initial reasoning is automatic, irrepressible, and hence unconscious. However, this interpretation is in no way obligatory, because there is no proof that Part 1 of the training instructions is in any way effective. It is also possible to explain performance improvement more prosaically, as a consequence of the fact that subjects have learned to apply the strategy given in Part 2 of the training instructions. To summarize, in the first phase subjects reason by manipulating transient and conscious mental models in a way that they believe to be correct, whereas after training they consciously apply the strategy that the experimenter declared to be correct. All of this is fully consistent with the SOC framework.

The Houdé et al. (2000) study includes positron emission tomography (PET) scan measures of cortical activities. Although it is not clear whether **Houdé** considers imaging



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

data as supplementary evidence for unconscious reasoning, it is worth emphasizing that they cannot be interpreted in this way. Neuroimaging techniques as such are obviously unable to show whether any given cortical activities are conscious or unconscious. If activities are a priori conceived of as unconscious, then cortical images will be said to show unconscious cortical activities, but this amounts to no more than restating the initial postulate. The pattern of cortical activations is also fairly consistent with our interpretation. Because conscious activities differ between the two phases of the experiences, they naturally correlate with the activation of different cortical areas. The activation of an emotion-related area in the production of responses after training, which Houdé interprets as "the emotional component of the brain's error correction device," can easily be accounted for by the anxiety-arousing nature of the training instructions, the first part of which focuses on a warning concerning the presence of a "trap." In our view, this activation is completely unrelated to Damasio's theory in the way Houdé argues.

**Barrouillet & Markovits** agree with us that the SOC model is more consistent with the theory of mental models than with the theory of mental logic. However, they also point out that we may have underestimated the residual points of divergence. For example, Johnson-Laird and Byrne (2002) assume that an abstract schema that they call the "core meaning" the content of which is roughly similar to the logical rules posited by the proponents of mental logic, is available to subjects performing conditional reasoning. This assumption indeed departs from the SOC principles insofar as such an abstract schema is both representational and unconscious. Markovits and Barrouillet (2002a), however, propose an alternative model that although still in keeping with the main tenets of mental models, obviates the need for a core meaning in conditional reasoning. Their reliance on the distributional properties of the specific events involved in the task makes the stance of their proposal very close to the SOC framework.

However, although Markovits and Barrouillet's model works well whenever the task involves concrete and familiar events, it is unable, according to **Barrouillet & Markovits**, to account for cases in which the premises are abstract and unfamiliar. Of course, the difficulty of reasoning in such cases is well documented, but recent evidence produced by Venet and Markovits (2001) suggests that the phenomenon exists, although rare. This would imply that the notion of "core meaning" cannot be totally dismissed. The interpretation Venet and Markovits (2001) suggest, which draws on the Representational Redescription (RR) mechanism advocated by Karmilov-Smith (1992), is irreconcilable with our framework. Indeed, RR assumes that abstract principles preexist in an implicit format before they are redescribed into a more explicit format (Vinter & Perruchet 1994). In other words, RR relies on a sophisticated cognitive unconscious. Is there an alternative explanation compatible with the SOC model?

### *R6.3. The SOC and the abstraction issue*

To address this question, let us recall how the SOC framework accounts for abstraction. This will allow us, at the same time, to address **Seger**'s criticism, according to which our framework does not differ in a principled way from the classical one in this regard. In the SOC model, abstraction is the by-product of mechanisms of strengthening and forgetting that naturally select the features common to a number of conscious representations, whereas the idiosyncratic features vanish. This explanation has a weak explanatory power insofar as the initial coding is stimulus-bound. However, the explanatory power is notably increased when the initial coding is allowed to be relational.

For the sake of illustration, let us imagine a conditioning situation in which S5 is positively reinforced and S3 nonreinforced, with 3 and 5 coding two levels of luminance of a stimulus S. If the animal is subsequently presented with S5 and S7, the animal (at least mammals) will choose S7, not S5. If S3, S5, and S7 are coded idiosyncratically according to their respective luminance, this result cannot be encompassed within a simple associative account. Indeed, we need a post hoc analogical reasoning mechanism that is capable of identifying that S7 is to S5 what S5 was to S3 and then inferring that the correct response is S7. However, let us imagine that the initial events are directly coded in a relational format through hardwired neural mechanisms (there is evidence that such a relational coding occurs as early as the retinal level). The reinforced stimulus is no longer S5, but the brightest of the two. A post hoc inferential step is no longer necessary: The feature "brighter than" can be subject to mechanisms of strengthening and forgetting in the same way as any other feature, thus allowing immediate transfer. We argue at length in the article (sect. 6) that it suffices to postulate that the initial coding provides information on simple relations, such as same/different, repetitions, or alternation, to account for the data on transfer available in the experimental literature. Contrary to **Seger**'s claim, this interpretation strikingly differs from the classical one. In keeping with the SOC's principles, the online formation of conscious representations is traded against sophisticated offline computations.

Although the possibility of relational coding substantially increases the explanatory power of our model, this power remains limited, hence allowing easy falsification. Do Venet and Markovits's results provide sufficient evidence to falsify our model? The fact that, in our model, abstraction emerges as the end result of an iterative mechanism appears consistent with **Barrouillet & Markovits**'s claim that such abilities are late-developing. However, to be encompassed within the SOC model, this hypothesis implies that the abstract feature to be selected remains simple enough to be processed during the life span of a particular conscious experience. Barrouillet & Markovits note that fulfilling this condition in accounting for Venet and Markovits's results appears problematic. We agree, but two remarks are in order. First, as for any isolated report, it appears necessary to confirm the empirical robustness of the phenomenon to be explained. Second and more important, further studies are needed to assess more precisely the explanatory scope of the SOC model in this regard.

Indeed, the phenomenon on which we base our approach here, namely, the direct coding of the incoming information in a relational format, has rarely been taken as an object of study in mainstream research. Of course, a notable exception is the work of James Gibson (e.g., 1966a). **Shanon** points out that for Gibson and his followers, "the units of perception (and of cognition at large) are not atomistic sensory inputs but higher-order meaningful relations." Gibson's theory of direct perception avoids introducing unnecessary mediating representational structures, and our


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

account of transfer and abstraction heavily relies on such ideas. However, the computational framework and its reliance on concepts such as rule-directed inference remain prevalent. As a consequence, very little is known about the power and limits of this mode of processing. A natural implication of the SOC is to stress the need for an examination of what actually occurs during a unit of psychological time. **Murray** provides us with an analysis of the effortless detection of a repetition among a set of elements concurrently held in consciousness. Further such analyses are needed in order to substantiate the conception of abstraction that we have outlined and to define what kinds of data are consistent with this theory and what kinds of data could invalidate it.

#### *R6.4. Incubation and creativity*

Our interpretation of problem-solving and incubation (sect. 7.1.2) relies heavily on the selection of relevant elements because of the forgetting of irrelevant ones. **Velmans** challenges this interpretation, arguing that it cannot account for a creative solution. He asks: "Can this really be explained in terms of selective forgetting of inappropriate associations? If creativity were solely a matter of retrieving appropriate, *existing* associations, what about creativity could possibly be *creative*?"

An analogy with the biological theory of evolution is helpful here. This theory is also grounded on a selection process that eliminates bad solutions. However, evolution is creative, in the sense that complex systems emerge from simpler ones. The reason is that there is a step before selection, namely, the existence of variability. Our response to **Velmans** is that he disregards this step in his reasoning. But the analogy with biological evolution, if it is to be compelling, needs to address at least two additional questions. First, in biological selection, the source of variability is well identified: it is the random mutation of the genetic material, followed by genetic mixing. What is the source of variability in the SOC framework? Second, in biological selection, the reason for selection is also clear-cut: survival. What makes it possible to select a correct solution in cognitive activities?

Regarding the first point, it should be noted that at an elementary level, PARSER contains a source of variability. Indeed, each attentional chunk comprises a random (and hence variable) number of perceptual primitives. In a real-life context, this source of variability, linked to the composition of the environmental features that are selected for entry into the current phenomenal experience, is exponentially amplified by the number of possible solutions offered by the richness and the multidimensionality of the environment. But we guess that this response remains unsatisfactory for **Velmans**, because whatever the number of selected primitives, the solution that emerges is embedded in the external data. This argument meets a criticism often leveled at association-based accounts, namely, their heavy reliance on data-driven processing. However, this criticism applies only to the outdated S-R theory, in which associations are formed between current events. In the target article, we emphasized the fact that modern associative learning theories hold that associations occur between representations (sect. 2.2.2). This conception frees association-based products from their exclusive dependence on the current sensory input, because the representations entering into associative links can also relate to past events. The source of variability thus becomes extraordinarily rich. Selection can now operate not only among the different ways of segmenting the current input, as in PARSER, but among the infinite combinations of externally and internally driven representations. In passing, this property also addresses **Dienes & Perner**'s objection, according to which we should even be unable to "account for thoughts about nonexisting entities, e.g., a unicorn galloping."

The analogy with biological evolution is limited, however. After all, it is nice to be able to account for our ability to think about a unicorn galloping, but in real life our conscious experiences are not continuously occupied by such flights of fancy. The evolutionary process of cumulative blind variation and selection does not dovetail easily with our actual experiences, except maybe our dreams. The reason is that representations are not generated through a blind process linking together any one primitive to any other primitive in random ways. Instead, the constitutive elements are generated as a function of their associative relevance. However, this consideration does not undermine our main argument: A selection-based mechanism can account for creativity because the initial source of variability is not limited to the current sensory input.

The second question raised by our analogy with biological evolution theory is far more difficult to address. What, in knowledge formation, plays the role equivalent to that of survival? In more specific terms, how are we able to know that a specific combination of perceptual and representational primitives within one attentional chunk makes sense, and provide the solution to a problem? The issue is reminiscent of the question that Meno asks of Socrates in Plato's dialogue: "If you find what you want, how will you ever know that this is the thing which you did not know?" (for an insightful discussion, see Cziko 1995, especially Chs. 6 and 11). We are afraid that we have no new insight to add to the centuries of philosophical debates. It is worth noting, however, that the classical computational framework is no better equipped, with its constant recourse to solutions based on the concept of innateness being little more than an avowal of ignorance.

### R7. Subliminal perception, dissociations, and related phenomena

#### *R7.1. Subliminal phenomena in normal people*

The phenomenon of unconscious semantic priming requires that the processing system uses a symbolic representation without any conscious counterpart, a requisite that is at odds with SOC's principles. In the target paper (sect. 8.2), we argued that the phenomenon is not empirically grounded. In order to complement our reference to earlier critical syntheses, we presented a fresh analysis of the Dehaene et al. (1998) study, in which the speed of classification of a target number as larger or smaller than five was affected by the congruency of a masked prime number. As pointed out by **Holender & Duscherer**, this result, typical of much of the recent evidence for unconscious semantic access, differs from the pure semantic priming effects alleged in older studies. Indeed, it can be shown that the effect is caused by response competition. If this effect occurs in subliminal conditions, then this entails that the prime is unconsciously identified, then unconsciously clas-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

sified on the basis of the same rule as the one applied to the visible target. Irrespective of the functional oddity of such a scenario, it is worth mentioning that it implies an unconscious that is even more powerful than the simple priming effect. In the article, we argued that the effect reported by Dehaene et al. was in fact neither unconscious nor semantic. Our reanalysis was not disputed by the commentators, but **Velmans** noted that not all the evidence can be as easily dismissed.

**Bornstein**, **Velmans**, and **Wolters & Phaf** mention a number of results that they see as incompatible with the denial of a cognitive unconscious. Although it is beyond the scope of this response to deal with the cited studies in detail, all the commentators adopted a similar line of argumentation that is worth discussing. As pointed out by Wolters & Phaf, the alleged evidence for semantic processing in the absence of conscious detection is not compelling, insofar as it can still be argued that conscious effects are not exhaustively excluded. For example, William James, **Shanon** reminds us, already emphasized the possibility of transient awareness as a factor undermining the interpretation of subliminal perception. However, there is a pattern of results that seems to escape such a limitation, namely, when the effects qualitatively differ according to whether the stimuli are above threshold or below threshold. A qualitative difference in performance is taken as an indication of a qualitative difference in processing, which is itself interpreted as evidence that the conscious/unconscious border has been jumped over.

We argue here that, again, the validity of this evidence is entirely dependent on the researcher's conception of mind. Within a conception in which there is a continuum between unconscious and conscious mental events, it is natural to think in terms of a trade-off, with expressions such as "less conscious" and "more unconscious" being used interchangeably, as illustrated in several commentaries (e.g., **Destrebecqz & Cleeremans, Wolters & Phaf**). When consciousness is degraded for any reason, the residual effects are naturally attributed to unconscious factors. However, this inference amounts to nothing more than the initial postulate, namely the existence of a continuum. If one postulates, as we do, that there is no unconscious mental life, the very same data becomes readily interpretable in other ways.

Our view is that dissociations between the effects of a given event do not distinguish between conscious and unconscious activities but between different conscious experiences. As we claimed earlier, "all the methods devised to show dissociations between conscious and unconscious phenomena actually contrast various forms of conscious experience, all of which are due to unconscious mechanisms" (Perruchet et al. 1997b). Let us consider how our account works for the question we are addressing here, through an examination of Jacoby's fame judgment experiments (e.g., Jacoby et al. 1989). In these experiments, subjects have first to read the names of nonfamous people in various conditions, then they are asked in a subsequent test to decide whether names, including those from the study list, are famous or not. When the study task is performed under optimal conditions, it obviously helps subjects perform the test. However, when the study phase occurs under impoverished conditions of presentation or divided attention, the opposite result is obtained: The names belonging to the study list are said to be famous in the subsequent test more often than in a control group that has not seen the study list. This result is conventionally interpreted as an unconscious effect of the prior exposure to these names. Our account of this result (Perruchet et al. 1997b; see also Mulligan & Hirshman 1997) is far simpler: Subjects explicitly remember that they have read these names in the past, but do not remember the relevant aspects of the spatiotemporal context of encoding, namely that they read these names in a list of nonfamous people. The false fame effect in Jacoby's experiments does not reveal the unconscious effect of an earlier episode but, more prosaically, false or impoverished conscious memory.

We believe that the line of reasoning followed above is readily generalizable. Thus, we conjecture that the results showing qualitatively different effects when the detectability of the stimuli is decreased can be accounted for by the effect of different conscious experiences, as well as by the postulate of a dual source of influence, respectively conscious and unconscious. In other words, we replace the conscious/unconscious axis with a multifaceted view of conscious experiences, in which conscious experiences may, as a function of the experimental conditions, differ in nature, clarity, associated contents (e.g., whether contextual elements are present), and so on. All these potential differences may account for differences in performance, eliminating the need for the cognitive unconscious postulate.

We do not claim that we have a full-blown alternative interpretation for all the experimental results put forth by the commentators. For example, we acknowledge that we fail to make sense of Groeger's (1998) results reported by **Velmans**. It should be noted, however, that endorsing the postulate of a cognitive unconscious does not help very much. One may wonder why, in Groeger's data, the physical form of the cue has no effect in subliminal conditions and likewise why the meaning of the cue has no effect in supraliminal conditions (even though the results in subliminal conditions suggest that semantic access is automatic). This result pattern completely contradicts, for example, the interpretation proposed by **Wolters & Phaf** to account for the reports of qualitatively different effects, namely, that "the main qualitative difference may be that unconscious processes are much less 'powerful' than conscious processes."

**Bornstein** comments on the subliminal mere exposure effect (SME), in which a preference judgment is enhanced by the presentation of a stimulus that cannot be subsequently recognized. He enumerates a number of properties that make this phenomenon worth considering, and notes that "SME effects are particularly problematic for the SOC model, because not only do stimuli perceived without awareness produce robust changes in affective responding, but these changes are significantly *greater* than those produced by identical stimuli that are consciously perceived" (see also **Wolters & Phaf**). Although the SME effect is an interesting phenomenon in itself, we consider that neither the judgment/recognition dissociation nor the enhancement of the SME with lowered conditions of presentation provides a compelling argument for claiming that a consciousness threshold has been crossed, for the reasons indicated above. Regarding Bornstein's hypothesis that unconscious effects are especially likely to emerge because the judgment is affect-based, two types of counterargument have been put forward in the literature, which we can allude to only briefly here. First, the same effect has been obtained with brightness and darkness ratings, which do not involve any affect (Mandler et al. 1987), and second, the dif-


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

ference between affect-based judgment and recognition has been recently reinterpreted as a difference in subjects' strategies (Whittlesea & Price 2001, see also **Holender & Duscherer**).[11]

We are fully aware that our handling of this issue is cursory and certainly insufficient to convince those who endorse a more classical view. To conclude this discussion, we would like to underline two aspects that may incite unconvinced readers to reconsider their positions. First, skepticism about the existence of subliminal semantic access is not as marginal an attitude as a reading of the textbooks, for example, might suggest. The classical interpretation is refuted by some firsthand researchers in the field. Holender led the way in his earlier *BBS* target paper (Holender 1986), and his present commentary adduces fresh analyses supporting the view "that the available data on unconscious semantic access do not constituted a challenge to the mentalistic framework" (**Holender & Duscherer**; see also **Dulany**). Second, even those who do not endorse the radical view we advocate here acknowledge that the effects are typically weak and short-lived. **Lambert** notes: "this cautions against extravagant theorizing concerning the functional role of unconscious semantic processes." We believe that Lambert's cautionary advice can be taken further. If the only reason behind so strong a postulate as an unconscious mental life turns out to be the barely detectable experimental phenomena that some researchers identify in their data, the time may have come to explore alternative theoretical avenues.

#### *R7.2. Neuropsychological dissociations*

**Lambert** states that: "the most cogent grounds for accepting that complex cognitive processes can be dissociated from consciousness comes not from experimental psychology, but from neuropsychology." In the target article, we devoted only a short passage (sect. 8.3) to neuropsychological dissociations. To our surprise, with the exception of Lambert, the commentaries that dealt with this issue tended to support our view. Of special interest is the fact that two of them (**Siéroff**, **Bartolomeo & Dalla Barba**) come from researchers who are also neuropsychologist practitioners.

We argued in the paper that it remains at issue to know whether neuropsychological disorders illustrate more than above-chance but degraded performance accompanied by a distorted but nevertheless present phenomenal awareness. Regarding the first part of the argument, namely, the fact that performance is degraded compared to normals, **Siéroff** extends our claim, which focused mainly on blindsight, to visual extinction. Visual extinction is a clinical sign of hemineglect, in which patients are unable to report about a stimulus displayed on the side opposite to their lesion whenever another stimulus is displayed on the other side and especially when the two stimuli are identical. However, patients still exhibit effects such as repetition and semantic priming when the prime is displayed on their "bad side." Siéroff emphasizes the fact that these residual abilities should not mask the deep behavioral impairments in such patients, together with their devastating clinical consequences.

Regarding the conscious experiences of patients, our coarse evocation of a "distorted but still present" phenomenal awareness provides an initial attempt at an interpretation. For example, Munakata (2001) uses the concept of graded representations to account for visual extinction, thus avoiding the postulate of a separable system for conscious awareness. However, several commentaries help us to refine our arguments by invoking more qualitative differences between experiences. **Dulany** and **Bartolomeo & Dalla Barba** show how the distinction between first-order (or direct) consciousness and second-order (or reflexive) consciousness (see sect. R2.1) allows us to reinterpret many alleged dissociations between conscious and unconscious forms of processing. Blindsight may be thought of as an example of preserved direct consciousness without reflexive consciousness, whereas Bartolomeo & Dalla Barba identify the capacity of some patients with hemineglect to acknowledge their symptoms without being able to remedy them as examples of the opposite dissociation. Dulany, on the other hand, interprets prosopagnosia as a dissociation between literal awareness of facial forms and identity awareness of the face, instead of revealing an intact recognition system disconnected from conscious awareness.

We agree with **Lambert** that all of these interpretations need to be strengthened and better articulated if we are to account for the full range of neuropsychological syndromes. However, we do not find any compelling counterarguments in his commentary. The main counterargument he puts forward is borrowed from Weiskrantz (1997), who claims that there are qualitatively different patterns of performance in blindsight and degraded normal performance. But we do not posit that degraded performance because of neurological lesions has to be like the degraded performance resulting from impoverished conditions of stimulus presentation in normal subjects. Shortsighted people do not see in the same way as normal-sighted people in the fog, although sight may be considered to be degraded in both cases. Simply stated, sight can be degraded along a multitude of dimensions. Why not consciousness? We side with **Bartolomeo & Dalla Barba** in concluding that "different varieties of consciousness may be involved in human behavior, before invoking 'unconscious' processes as a default explanation for dissociations."

### R8. A radical option?

Although a few commentators express strong and definitive aversion to the general stance of the target article (e.g., **Litman & Reber**), many others express at least some sympathy. However, most of those who see at least some value in our attempt find it too radical. **Sun** states that "we should be careful not to allow limited results from our favorite paradigm to shape our entire world view. Overgeneralizing useful, but limited, results is methodologically harmful." To end this response, we argue first that our framework, despite its large scope, is nevertheless not intended to be all-encompassing. Second, we outline the reasons motivating the defense of a radical proposal in the area on which we concentrate.

#### *R8.1. From conscious content to overt behavior*

We have emphasized above (R1.2) that our research objective was to account for the content of the phenomenal experience and not directly for overt behavior. However, it is also true that in some places in the paper we referred in-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

cautiously to "adaptive behavior" or used similar general expressions, hence causing potential confusion. We believe it necessary to note that our paper is incomplete in two respects. First, concerning behavior that seems highly dependent on conscious knowledge, our target article says nothing about the route that leads from phenomenal consciousness to overt performance. We do not embark on this discussion here, given that the point was not alluded to by commentators. It does, however, deserve investigation. A second problematic aspect concerns the problem of knowing whether any adaptive behavior is causally related to phenomenal consciousness.

For example, what about the case described by **Keisler & Willingham**? These commentators mention studies illustrating our ability to compensate for an external force field during a motor task. "The motor system," they write:

> computes the appropriate counterforce from knowledge of (1) the external force and (2) properties such as velocity, position, orientation, and so forth, of the hand. The fact that the system must *extrapolate* the appropriate counterforce based on this information is the critical evidence of unconscious "high-level conceptual processing," against which P&V argue.

**Keisler & Willingham** rightly point out that some characteristics of the phenomenon do not fit well with an association-based view. Moreover, the interpretation they suggest, which involves computations requiring knowledge about the biomechanical properties of the hand as well as about the properties of the external force, is obviously at odds with the SOC framework. Prima facie, such an example of skill-learning (see also **Dienes & Perner**) constitutes a strong challenge to our position. Our suggestion is that our model is indeed irrelevant for explaining motor adjustments but that Keisler & Willingham's interpretation is not compelling either. In other words, we believe that this form of motor learning could depend neither on phenomenal experience, nor on the unconscious manipulation of unconscious representations.

An alternative explanation was proposed long ago in terms of dynamic systems. The dynamic system theory (e.g., Kelso 1990; Kelso & Schöner 1988) focuses on the evolution of a system over time as a function of a set of variables that interact in accordance with specific principles. It has been applied with success to describe various motor adjustments to changes in the physical environment. "Dynamic system" explanations of behavior share many features with the SOC framework: They focus on learning, they are centered around self-organizing principles guiding dynamic and time-locked interactions between the organism and its environment, and as a consequence, they obviate the need for a cognitive unconscious involving knowledge manipulation, as in **Keisler & Willingham**'s interpretation. However, they fundamentally differ from our proposal because they exclude any mental representations, whether conscious or unconscious. In our opinion, this property makes them inappropriate for accounting for human behavior outside specific domains. But the same property makes them quite attractive in the domains in which they can be applied. In some sense, relying on dynamic system theory to account for some instances of skill-learning echoes one of the major arguments used by William James against the postulate of unconscious mental structures. As **Shanon** reminds us, William James operated a fundamental distinction between "brain facts" and "mind facts" and argued that, to quote Shanon:

> If psychological phenomena can be directly explained in terms pertaining to the brain, no additional covert psychological structures and processes have to be invoked in order to account for them.[12]

An appealing speculation is that in the near future, the SOC model and dynamic system theory might be found to rely on self-organizational principles that are basically similar but applied to different levels of description and explanation.

### *R8.2. Is such a radical option warranted?*

The above restriction of the area of relevance of the SOC framework does not eliminate the fact that it constitutes a radical option for a very large range of applications. **Ellis** rightly notes that we "force a choice between extreme viewpoints, and ignore much middle ground." Several commentators advocate moderation on the grounds that a scientific approach must secure itself against premature generalizations. Representative of this attitude, as we read it, is **Redington**'s commentary. Redington's recommendation to avoid overgeneralization is worthy of special interest, given that his own studies have greatly contributed to showing the power of a distributional approach in accounting for performances in various domains from implicit learning to syntactic competences.

To a very large extent, we are ready to acknowledge that the SOC model is overly speculative. Empirical data are certainly insufficient to provide definitive evidence for many of our arguments. As a consequence, we understand that our radicalism may be thought of as a lack of scientific rigor, as **Sun** and others argue. However, we believe that in the present context, the choice of a radical position is the best one. Below, we substantiate the claim that we have arrived at a juncture where the risk of overgeneralization is outweighed by the potential heuristic value of a drastic change in our vision of mind.

Our main argument stems from an examination, not of the current state of the art, but of the recent trends in research. Two major trends are worth considering. The first concerns consciousness. During the last two decades, empirical supports for the analytic capabilities of unconscious cognition have gradually waned. As a case in point, in the domain of subliminal effects, Greenwald and coworkers, after having examined the possibility of analyzing the meaning of a two-word sentence (Greenwald 1992), now reject the possibility of analyzing even a single word and examine the possibility of analyzing the meaning of parts of a word (Abrams & Greenwald 2000). A similar trend is apparent in the areas of implicit memory and learning. At the same time, interest in consciousness has been growing, as testified to by the recent upsurge of journals and books devoted to this issue. A second trend in research concerns the gradual decline of the role attributed to rules in the architecture of the mind, which have been giving way to interpretations based on mnesic or statistical processes. The same trend is apparent in virtually all domains of research. For example, the emblematic debate on the past tense in English seems to be virtually over to the detriment of rule-based approaches (see Ramscar 2002). Even in cases where a simple rule applies consistently, as for the spelling of the plural endings in English (all regular plurals are spelled with -s), both children's and adults' behavior is now explained as a sensitivity to the distributional features of the input (Kemp & Bryant 2003).


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Our proposal is situated at the meeting point of these two nearly independent developments. By suggesting that consciousness, through its self-organizing property, naturally makes us sensitive to the distributional structure of the world, hence accounting for behavioral adaptation, we simply push these trends to their ultimate ends while also providing a meaningful story for their integration.

It could be argued that if the growth of ideas compatible with our framework is as patent as we contend, it would be methodologically safer to await their natural emergence than to make premature speculations. But this commonsense approach suffers from a paradox. Let us consider the evolution of psychokinesis, the possibility of moving objects through the power of the mind. Around 1850, the objects in question were heavy tables of around 100 kilograms. A few decades later, tables were abandoned in favor of saucepans and other kitchen utensils weighing about 1 kilograms. Around 1970, only small objects such as chess pieces were involved. Now the only claims concern the ability to move a tiny sheet of paper weighting about 1 gram (Charpak & Broch 2002). However, belief in psychokinesis has not collapsed in line with the decline in the supporting evidence that has resulted from more stringent controls. The reason is that the claims that remain become both less extravagant and increasingly difficult to refute. The very same story may happen for unconscious semantic access or unconscious rule abstraction. Anyone wishing to keep alive the postulate of a cognitive unconscious can indeed find some evidence for it in the literature, irrespective of how tiny and transient this evidence becomes, and this situation may persist forever.

Exploring a new path before forming a comprehensive refutation of the arguments supporting the classical framework is heuristic, we argue, because it opens new avenues of research. In section R5.2, we pointed out that the argument of the "poverty of stimulus" put forth by psycholinguists of Chomskyan persuasion to dismiss empiricist approaches is valid only if one subscribes to a Chomskyan conception of mind. The input may be too poor to support rule extraction but rich enough to support other forms of learning: All depends on how the mind works. Our own claim is that learning is a by-product of conscious experiences. Up to now, we have assumed that consciousness plays the role of a statistical analysis device such as a connectionist network. But this is only a coarse-grained approximation. Indeed, the input to a connectionist network commonly comprises only the current sensory information, whereas we have argued at length that in the SOC framework, associations occur between complex representations that are not necessarily linked to the displayed material. As a consequence, the proposal that the distributional information embedded in the input may not be sufficient, as **Redington** notes, could itself be linked to a conception in which learning proceeds through statistical analysis of the input, in the same way that the classical "poverty of the stimulus" argument is linked to a conception in which the mind works through hypothesis testing and rule induction.

Here is our main response to the advice offered by **Redington** and others who advocate a middle way: To know whether a new framework is valid, we have not only to search for evidence in support of it within the data resulting from studies carried out within the current framework, but we have also to furnish new evidence. This entails making a bet. We hope to have shown that the bet we propose is reasonable, but we are aware that only future research can determine whether it is correct.

ACKNOWLEDGMENTS

This research was supported by the Centre National de la Recherche Scientifique (UMR CNRS 5022), the University of Bourgogne, and the Region of Bourgogne.

NOTES

**1. Litman & Reber** wrote:

> The question of whether the unconscious follows rules that are instantiated symbolically or whether it functions on associative principles is, despite the intense focus P&V put on it, actually irrelevant. The fundamental questions really under the microscope here, are the complexity of the problems the unconscious can solve and the nature of the tasks the unconscious can perform. The particular strategy that evolved to solve those complex problems is a secondary problem.

This commentary proceeds from a logical flaw. Indeed, it is impossible to know whether a problem has been solved consciously or unconsciously without knowing the particular strategy that has been used. For example, Reber has long attributed improved performance in artificial grammar-learning settings to unconscious abstraction (e.g., Reber 1989). This assessment was based on the observation that subjects were unaware of the rules of the grammar. However, subsequent research demonstrated that subjects did not rely on rules but learned small fragments of the material. When researchers' attention was drawn to this knowledge content, it was seen to be conscious (e.g., Perruchet & Pacteau 1990). This illustrates the somewhat trivial assertion that before we can know whether a form of processing is conscious or unconscious, we must identify, at least roughly, this form of processing.

**2.** We use "world" as a shortcut to designate whatever interacts with the mind, whether material or not, permanent or transient, inside or outside one's organism, and so on. As a case in point, language is included.

**3. Parisse & Cohen** rightly note that PARSER deals with an artificial material that is not representative of natural spoken language. In particular, syllables always have the same physical instantiations, hence making their identification easier than in real speech. We agree that PARSER is not a full-blown model of word segmentation. Its primary objective is to show that the SOC concept can perform the task that is usually attributed to symbolic computations or to the analysis of the distributional properties of the input. Parisse & Cohen's observation does not undermine this comparison, because the raw material used in the other relevant studies is similar.

**4.** Although we see nothing here other than terminological issues, **Dienes & Perner** go on to assert that our position "would amount to the . . . claim that if a representation is fact-explicit, it will automatically be attitude-explicit." We fail to see the logic of their argument here, except if their own terminological option, namely that a fact must be attitude-explicit to be ascribed a conscious status, is given a normative status instead of remaining a purely terminological proposition.

**5.** The word "computation" is also potentially confusing. Throughout the target article and this response, in line with common practice, we use terms such as "computational," "classical," or "cognitive" interchangeably to designate the framework with which we contrast our model. On the other hand, we ourselves rely on a computational model, PARSER. To be clear, when we argue against computationalism, our target is the prevalent conception according to which the brain performs unconscious operations in some way analogous to those performed by a digital computer (and consequently analogous to conscious operations). Of course, we acknowledge the validity of computational models as simulations of unconscious processes, but in the very same way that we acknowledge the ability of computational models to sim-



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

ulate "weather or planetary movements" (**Lambert**). The subtractions performed in PARSER to simulate forgetting, for example, mimic the effect of processes that have strictly no relation to an arithmetic operation performed on symbolic tokens. On the other hand, we also acknowledge the possibility of conscious computations and the possibility of simulating those computations with a computer program in a real sense.

**6.** The potential circularity of our approach is nicely summarized by **Seger** in the following formula: "Is it conscious? must be part of the representation. Is it unconscious? must be part of the processes working on the representation." However, the argument is a two-edged one. If there is no way to dissociate processes and representations, the claim for unconscious representations is also unfalsifiable.

**7.** In the developmental literature, the meaning of "mental" is somewhat restricted, essentially owing to the writings of Piaget (1952). In his description of developmental stages, Piaget assimilated "mental" with "conceptual" and contrasted conceptual and sensory-motor stages. This leads him to consider infants to be devoid of any mental activity (in this strict sense). We do not refer to this specific meaning here.

**8. Lambert** wrote: "The challenge facing cognitive science is to understand how mental phenomena, including consciousness, can arise from the collective activity of millions of neural components, each of which may be viewed as performing computational operations." We believe this endeavor to be inherently flawed. No science addresses the challenge of understanding a football match on the basis of the activity of billions of physical particles. Instead, we attempt to understand mental phenomena in terms of self-organization, that is to say, after having selected the mental as our level of description and explanation.

**9. Destrebecqz & Cleeremans** seem to agree with this contention when they claim that "one could choose to describe [the weak instances of graded representations] as part of a 'fringe' of consciousness as unconscious representations, or transient representations." However, in so doing, they themselves undermine both their criticism of our framework and the basis for their alternative view (e.g., Cleeremans & Jiménez 2002).

**10.** For example, because syntactic knowledge requires that words are individuated, a logical approach should have consisted in first investigating how words are extracted from continuous speech. The reason for the actual story to be different is that word segmentation can hardly be grasped by rule-based, innatist approaches.

**11.** Because our response to **Bornstein** could be viewed as a denial of the value of taking affect into account in psychological theorizing, we feel it necessary to emphasize that our position lies at the opposite extreme. We acknowledge with Bornstein that, in keeping with the current literature, our analysis "is limited primarily to studies that use affectively neutral stimuli and outcome measures." However, one major advantage of the SOC, in our view, is that it make it possible to reintegrate affective factors within a full-blown conception of mind. Affective factors have no place in the classical information-processing, computational framework, because they are irrelevant for a computer. By contrast, our focus on phenomenal experience offers a natural opportunity to consider affective valence, because most conscious contents are affectively valenced (whereas the idea of an affect that would have no conscious manifestation seems singularly awkward).

Our reliance on associative mechanisms also helps guarantee that a place will be found for affective dimension. To do justice to the otherwise outdated behaviorist doctrine, researchers into learning long ago acknowledged the need for affect-related concepts, such as the concept of "drive." These possible extensions of the SOC framework were (too briefly) outlined in the target article (sect. 7.2).

**12.** This position could bear some similarity with the proposal made by **Wolters & Phaf** toward the end of their commentary, as we understand it. However, in contrast to Wolters & Phaf, we see the fact that some behavioral adaptations cannot be ascribed to consciousness as compatible with our framework, inasmuch as they do not involve mental activities.

## References

**Letters "a" and "r" appearing before authors' initials refer to target article and response, respectively.**


Abrams, R. L. & Greenwald, A. G. (2000) Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science* 11:118–24. [DH, rPP]

Abrams, R. L., Klinger, M. R. & Greenwald, A. G. (2002) Subliminal words activate semantic categories (not automated motor responses). *Psychonomic Bulletin and Review* 9:100–106. [DH, aPP]

Adolph, K. E. (2000) Specificity of learning: Why infants fall over a veritable cliff. *Psychological Science* 11: 290–95. [aPP]

Allport, A. (1988) What concept of consciousness? In: *Consciousness in contemporary science,* ed. A. J. Marcel & E. Bisiach. Oxford University Press. [PBart]

Allport, D. A., Tipper, S. P. & Chmiel, N. R. J. (1985) Perceptual integration and postcategorical filtering. In: *Attention and performance XI,* ed. M. I. Posner & O. S. M. Marin. Erlbaum. [DH]

Altmann, G. & Dienes, Z. (1999) Rule learning by seven-month-old infants and neural networks. *Science* 284: 875. [aPP]

Anderson, J. R. (1978) Arguments concerning representations for mental imagery. *Psychological Review* 85:249–77. [CAS]

(1994) *Rules of the mind.* Erlbaum. [aPP]

Anderson, J. R. & Milson, R. (1989) Human memory: An adaptive perspective. *Psychological Review* 96: 703–19. [aPP]

Aslin, R. N., Woodward, J. Z., LaMendola, N. P. & Bever, T. G. (1996) Models of word segmentation in fluent maternal speech to infants. In: *Signal to syntax,* ed. J. L. Morgan & K. Demuth. Erlbaum. [aPP]

Aurell, C. G. (1983) Perception: A model comprising two modes of consciousness. Addendum: Evidence based on event-related potentials and brain lesions. *Perceptual and Motor Skills* 56:211–20. [RDE]

(1984) Perception: A model comprising two modes of consciousness. Addendum II: Emotion incorporated. *Perceptual and Motor Skills* 59:180–82. [RDE]

(1989) Man's triune conscious mind. *Perceptual and Motor Skills* 68:747–54. [RDE]

Auroux, S. (1994) *La révolution technologique de la grammaticalisation.* Mardaga. [CP]

Baars, B. J. (1995) Psychology in a world of sentient, self-knowing beings: A modest Utopian fantasy. In: *Mind and brain in the 21st century,* ed. R. L. Solso. MIT Press. [aPP]

(1997) *In the theater of consciousness: The workspace of the mind.* Oxford University Press. [aPP]

(1998) Metaphors of consciousness and attention in the brain. *Trends in Neurosciences* 21:51–89. [aPP]

Baars, B. J. & McGovern, K. (1993) Does philosophy help or hinder scientific work on consciousness? *Consciousness and Cognition* 2:18–27. [GW]

Baeyens, F., Eelen, P., Crombez, G. & Van Den Bergh, O. (1992) Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style, and contingency awareness. *Behavior Research and Therapy* 30:133–42. [aPP]

Bar, M. & Biederman, I. (1998) Subliminal visual priming. *Psychological Science* 9:464–69. [RFB, aPP]

Bargh, J. (1992) The ecology of automaticity: Towards establishing the conditions needed to produce automatic processing effect. *American Journal of Psychology* 105:181–99. [JT]

Bartolomeo, P. & Chokron, S. (2002) Orienting of attention in left unilateral neglect. *Neuroscience and Biobehavioral Reviews* 26(2):217–34. [PBart]

Baylis, G. C., Driver, J. & Rafal, R. D. (1993) Visual extinction and stimulus repetition. *Journal of Cognitive Neuroscience* 5:453–66. [ES]

Baylis, G. C., Gore, C. L., Rodriguez, P. D. & Shisler, R. J. (2001) Visual extinction and awareness: The importance of binding dorsal and ventral pathways. *Visual Cognition* 8:359–79. [ES]

Bechara, A., Damasio, H., Tranel, D. & Damasio, A. (1997) Deciding advantageously before knowing the advantageous strategy. *Science* 275:1293–95. [OH]

Bedford, F. L. (1989) Constraints on learning new mappings between perceptual dimensions. *Journal of Experimental Psychology: Human Perception and Performance* 15(2): 232–48. [ASK]




https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

(1993) Perceptual and cognitive spatial learning. *Journal of Experimental Psychology: Human Perception and Performance* 19(3):517–30. [ASK]
Berry, D. C. & Broadbent, D. E. (1988) Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology* 79:251–72. [aPP]
Berry, D. C. & Dienes, Z. (1993) *Implicit learning: Theoretical and empirical issues.* Erlbaum. [aPP]
Besner, D. (2001) The myth of ballistic processing: Evidence from Stroop's paradigm. *Psychonomic Bulletin and Review* 8:324–30. [aPP]
Besner, D, Stolz, J. A. & Boutilier, C. (1997) The Stroop effect and the myth of automaticity. *Psychonomic Bulletin and Review* 4:221–25. [aPP, JT]
Bickhard, M. H. (1993) Representational content in humans and machines. *Journal of Experimental and Theoretical Artificial Intelligence* 5:285–333. [RDE]
Bjorklund, D. F. (1997) In search of a metatheory for cognitive development (or, Piaget is dead and I don't feel so good). *Child Development* 68:144–48. [OH]
Blake, R. (1998) What can be "perceived" in the absence of visual awareness? *Psychological Science X*:157–62. [aPP]
Bolinger, D. L. (1961) Syntactic blends and other matters. *Language* 37:366–81. [WS-YW]
Bonnano, G. A. & Stillings, N. A. (1986) Preference, familiarity, and recognition after repeated brief exposures to random geometric shapes. *American Journal of Psychology* 99:403–15. [RFB]
Bornstein, M. H. & Krinsky, S. J. (1985) Perception of symmetry in infancy: The salience of vertical symmetry and the perception of pattern wholes. *Journal of Experimental Child Psychology* 39:1–19. [aPP]
Bornstein, R. F. (1989) Exposure and affect: Overview and meta-analysis of research. 1968–1987. *Psychological Bulletin* 106:265–89. [RFB]
(1992) Subliminal mere exposure effects. In: *Perception without awareness,* ed. R. F. Bornstein & T. S. Pittman. Guilford Press. [RFB]
Bornstein, R. F. & D'Agostino, P. R. (1992) Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology* 63:545–52. [RFB]
Bower, T. G. R. (1979) *Human development.* Freeman. [aPP]
Braine, M. D. S. (1978) On the relationship between the natural logic of reasoning and standard logic. *Psychological Review* 85:1–21. [aPP]
Braine, M. D. S. & O'Brien, D. P. (1998) *Mental logic.* Erlbaum. [PB]
Brent, M. R. (1996) Advances in the computational study of language acquisition. *Cognition* 61:1–38. [arPP]
Brent, M. R. & Cartwright, T. A. (1996) Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61:93–125. [aPP, MR]
Brentano, F. (1774/1973) *Psychology from an empirical standpoint.* Routledge & Kegan Paul. [DED, RM]
Brewer, W. F. (1974) There is no convincing evidence for operant or classical conditioning in adult humans. In: *Cognition and the symbolic processes,* ed. W. B. Weimer & D. S. Palermo. Erlbaum. [RAC]
Broadbent, D. E. (1977) Levels, hierarchies and the focus of control. *Quarterly Journal of Experimental Psychology* 29:181–201. [aPP]
Bronson, G. W. (1982) *The scanning patterns of human infants: Implications for visual learning.* Ablex. [aPP]
Brooks, L. R. (1978) Nonanalytic concept formation and memory for instances. In: *Cognition and categorization,* ed. E. Rosch & B. B. Lloyd. Erlbaum. [aPP]
Brooks, L. R. & Vokey, J. R. (1991) Abstract analogies and abstracted grammars: A comment on Reber and Mathews et al. *Journal of Experimental Psychology: General* 120:316–23. [aPP]
Brown, M. S., Roberts, M. A. & Besner, D. (2001) Semantic processing in visual word recognition: Activation blocking and domain specificity. *Psychonomic Bulletin and Review* 8:778–84. [aPP]
Bruner, J. S., Goodnow, J. J. & Austin, G. A. (1956) *A study of thinking.* Science Editions. [CAS]
Brunn, J. L. & Farah, M. J. (1991) The relationship between spatial attention and reading: Evidence from the neglect syndrome. *Cognitive Neuropsychology* 8:59–75. [ES]
Bryant, P., Nunes, T. & Snaith, R. (2000) Children learn an untaught rule of spelling. *Nature* 403:157–58. [aPP]
Buchner, A. & Erdfelder, E. (1996) On assumptions of, relations between, and evaluations of some process dissociation measurement models. *Consciousness and cognition* 5:581–94. [aPP]
Bush, G., Luu, P. & Posner, M. I. (2000) Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences* 4:215–22. [OH]
Butler, L. T. & Berry, D. C. (2001) Implicit memory: Intention and awareness revisited. *Trends in Cognitive Sciences* 5:192–97. [aPP]
Campion, J., Latto, R. & Smith, Y. M. (1983) Is blindsight an effect of scattered light, spared cortex, and near-threshold vision? *Behavioral and Brain Sciences* 6:423–86. [aPP]
Cantagallo, A. & Della Sala, S. (1998) Preserved insight in an artist with extrapersonal spatial neglect. *Cortex* 34(2):163–89. [PBart]
Capalbo, M., den Dulk, P. & Phaf, R. H. (2000) A dual-pathway model for affective priming. *Journal of Cognitive Neuroscience* 61B:S54. [GW]
Carey, S. & Bartlett, E. (1978) Acquiring a single new word. *Proceedings of the Stanford Child Language Conference. Language Development* 15:17–29. [rPP]
Carlson, R. A. (1997) *Experienced cognition.* Erlbaum. [RAC]
(2002) Conscious intentions in the control of skilled mental activity. In: *The psychology of learning and motivation, vol. 41,* ed. B. Ross. Academic Press. [RAC]
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. [aPP]
Carlson, R. A. & Dulany, D. E. (1988) Diagnostic reasoning with circumstantial evidence. *Cognitive Psychology* 20:463–92. [DED]
Carr, T. H. & Dagenbach, D. (1990) Semantic priming and repetition priming from masked words: Evidence for a center-surround attentional mechanism in perceptual recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16:341–50. [DH]
Carruthers, P. (2000) *Phenomenal consciousness naturally.* Cambridge University Press. [ZD]
Case, R. (1993) Theories of learning and theories of development. *Educational Psychologist* 28:219–33. [aPP]
Catrambone, R. & Holyoak, K. J. (1985) Contextual limitations on analogical transfer. Paper presented at the 93rd Annual Convention of the American Psychological Association, Los Angeles, CA. [aPP]
Cavanagh, J. P. (1972) Relation between the immediate memory span and the memory search rate. *Psychological Review* 79:525–30. [DJM]
Charpak, G. & Broch, H. (2002) *Devenez sorciers, devenez savants.* Odile Jacob. [rPP]
Chater, N. & Oaksford, M. (1999) Ten years of the rational analysis of cognition. *Trends in Cognitive Sciences* 3: 57–65. [aPP]
Cheng, P. W. & Holyoak, K. J. (1985) Pragmatic reasoning schemas. *Cognitive Psychology* 17: 391–16. [aPP]
Cheesman, J. & Merikle, P. M. (1986) Distinguishing conscious from unconscious perceptual process. *Canadian Journal of Psychology* 40:343–67. [aPP, GW]
Christiansen, M. H. (1994) Infinite language finite minds: Connectionism learning and linguistic structure. Unpublished doctoral dissertation, Centre for Cognitive Science, University of Edinburgh, U.K. [WS-YW]
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. [aPP]
Christiansen, M. H., Conway, C. M. & Curtin, S. (2000) A connectionist single-mechanism account of rule-like behavior in infancy. In: *Proceedings of the 22nd Annual Conference of the Cognitive Science Society.* Erlbaum. [aPP]
Clark, A. & Thornton, C. (1997) Trading spaces: Computation, representation and the limits of uninformed learning. *Behavioral and Brain Sciences* 20:57–90. [aPP]
Cleeremans, A. (1993) *Mechanisms of implicit learning: A connectionist model of sequence processing.* MIT Press/Bradford Books. [aPP]
(1997) Principles for implicit learning. In: *How implicit is implicit knowledge?* ed. D. Berry. Oxford University Press. [AD]
Cleeremans, A., Destrebecqz, A. & Boyer, M. (1998) Implicit learning: News from the front. *Trends in Cognitive Sciences* 2:406–16. [aPP]
Cleeremans, A. & French, R. M. (1996) From chicken squawking to cognition: Levels of description and the computational approach in psychology. *Psychologica Belgica* 36:5–29. [LJ]
Cleeremans, A. & Jiménez, L. (1998) Implicit sequence learning: The truth is in the details. In: *Handbook of implicit learning,* ed. M. Stadler & P. Frensch. Sage. [aPP]
(2002) Implicit learning and consciousness: A graded, dynamic perspective. In: *Implicit learning and consciousness: An empirical, computational and philosophical consensus in the making,* ed. R. M. French & A. Cleeremans. Psychology Press. [AD, RMF, LJ, rPP]
Cleeremans, A. & McClelland, J. L. (1991) Learning the structure of event sequences. *Journal of Experimental Psychology: General* 120:235–53. [aPP]
Clement, C. (1994) Effect of structural embedding on analogical transfer: Manifest versus latent analogs. *American Journal of Psychology* 107:1–38. [aPP]
Cohen, A., Ivry, R. I. & Keele, S. W. (1990) Attention and structure in sequence learning, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16:17–30. [aPP]
Conditt, M. A., Gandolfo, F. & Mussa-Ivaldi, F. A. (1997) The motor system does not learn the dynamics of the arm by rote memorization of past experience. *Journal of Neurophysiology* 78:554–60. [ASK]
Cowan, N. (1995) *Attention and memory: An integrated framework.* Oxford University Press. [aPP]
(2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24:87–114. [aPP]



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Crabb, B. T. & Dark, V. (1999) Perceptual implicit memory requires attentional encoding. *Memory and Cognition* 27:267–75. [aPP]
Craik, F. I. M. & Lockhart, R. S. (1972) Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11:671–84. [aPP]
Crannell, C. W. & Parrish, J. M. (1957) A comparison of immediate memory span of digits, letters, and words. *Journal of Psychology* 44:319–27. [DJM]
Croft, W. (2001) *Radical construction grammar.* Oxford University Press. [CP]
Cziko, G. (1995) *Without miracles: Universal selection theory and the second Darwinian revolution.* A Bradford Book/MIT Press. [rPP]
Dagenbach, D., Carr, T. H. & Wilhelmsen, A. (1989) Task-induced strategies and near-threshold priming: Conscious influences on unconscious perception. *Journal of Memory and Language* 28:412–43. [DH]
Dalla Barba, G. (2002) *Memory, consciousness and temporality.* Kluver Academic. [PBart]
Dalla Barba, G., Bartolomeo, P., Ergis, A. M., Boissé, M. F. & Bachoud-Lévi, A. C. (1999) Awareness of anosognosia following head trauma. *Neurocase* 5(1):59–67. [PBart]
Damasio, A. R. (1994) *Descartes' error: Emotion, reason, and the human brain.* Grosset, Putnam. [OH]
(1999) *The feeling of what happens: Body and emotion in the making of consciousness.* Harcourt Brace. [OH]
Damasio, H., Grabowski, T., Frank, R., Galaburda, A. & Damasio, A. R. (1994) The return of Phineas Gage: Clues about the brain from the skull of a famous patient. *Science* 264:1102–105. [OH]
Damian, M. F. (2001) Congruity effects evoked by subliminally presented primes: Automaticity rather than semantic processing. *Journal of Experimental Psychology: Human Perception and Performance* 27:154–65. [DH, aPP]
Deacon, T. W. (1997) *The symbolic species: The co-evolution of language and the brain.* Norton. [WS-YW]
De Boer, B. (2001) *The origins of vowel systems.* Oxford University Press. [WS-YW]
Decaix, C., Siéroff, E. & Bartolomeo, P. (in press) How voluntary is 'voluntary' orienting of attention? *Cortex.* [PBart]
De Fockert, J. W., Rees, G., Frith, C. D. & Lavie, N. (2001) The role of working memory in visual selective attention. *Science* 291:1803–806. [GW]
De Haan, E. H. F., Young, A. & Newcomb, F. (1987) Face recognition without awareness. *Cognitive Neuropsychology* 4:385–414. [DED]
Dehaene, S., Naccache, L., Le Clec'h, G., Koechlin, E, Mueller, M., Dehaene-Lambertz, G, van de Moortele, P.-F. & Le Bihan, D. (1998) Imaging unconscious semantic priming. *Nature* 395:597–600. [DH, arPP]
De Houver, J., Hendrickx, H. & Baeyens, F. (1997) Evaluative learning with "subliminally" presented stimuli. *Consciousness and Cognition* 6:87–107. [aPP]
Dennett, D. C. (1982) Styles of mental representation. Proceedings of the Aristotelian Society 83:213–26. [rPP]
(1987) *The intentional stance.* MIT Press. [RDE]
(1991) *Consciousness explained.* Little, Brown/Bradford, MIT Press. [RM, rPP]
Dennett, D. C. & Kinsbourne, M. (1992) Time and the observer: The where and when of consciousness. *Behavioral and Brain Sciences* 15:183–247. [AJL]
Descartes, R. (1628/1961) *Rules for the direction of the mind.* Bobbs-Merrill. [OH]
Destrebecqz, A. & Cleeremans, A. (2001) Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin and Review* 8:343–50. [AD]
Diener, D. (1988) Absence of the set-size effect in memory-search tests in the absence of a pre-probed delay. *Memory and Cognition* 16:367–76. [DJM]
Dienes, Z. & Altmann, G. (1997) Transfer of implicit knowledge across domains: How implicit and how abstract? In: *How implicit is implicit learning?,* ed. D. Berry Oxford University Press. [aPP]
Dienes, Z. & McLeod, P. (1993) How to catch a cricket ball. *Perception* 22:1427–39. [ASK]
Dienes, Z. & Perner, J. (1999) A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences* 22:735–55. [ZD, rPP]
(2002a) A theory of the implicit nature of implicit learning. In: *Implicit learning and consciousness: An empirical, philosophical, and computational consensus in the making?,* ed. R. M. French & A. Cleeremans. Psychology Press. [ZD]
(2002b) The metacognitive implications of the implicit–explicit distinction. In: *Metacognition: Process, function, and use,* ed. P. Chambres, M. Izaute & P.-J. Marescaux. Kluwer Academic. [ZD]
Dishon-Berkovits, M. & Algom, D. (2000) The Stroop effect: It is not the robust phenomenon that you have thought it to be. *Memory and Cognition* 28:1437–49. [aPP]
Dixon, N. F. (1981) *Preconscious processing.* Wiley. [MV]
Draine, S. C. & Greenwald, A. G. (1998) Replicable unconscious semantic priming. *Journal of Experimental Psychology: General* 127:286–303. [DH, aPP]
Dretske, F. I. (1981) *Knowledge and the flow of information.* The MIT Press. [RM]
(1988) *Explaining behavior: Reasons in a world of causes.* MIT press. [LJ]
(1995) *Naturalizing the mind.* The MIT Press. [ZD, RM]
Dulany, D. E. (1968) Awareness, rules, and propositional control: A confrontation with S-R behavior theory. In: *Verbal behavior and general behavior theory,* ed. T. Dixon & D. Horton. Prentice-Hall. [RAC]
(1991) Conscious representation and thought systems. In: *Advances in social cognition, vol. 4,* ed. R. S. Wyer & T. K. Srull. Erlbaum. [DED, arPP, JT, GW]
(1997) Consciousness in the explicit (deliberative) and implicit (evocative). In: *Scientific approaches to the study of consciousness,* ed. J. D. Cohen & J. W. Schooler. Erlbaum. [PBart, DED, DH, AJL, RM, arPP, JT, GW]
(1999) Consciousness, connectionism, and intentionality. *Behavioral and Brain Sciences* 22:154–55. [aPP]
(2000) A mentalistic view of conscious unity and dissociation. Paper read at Association for the Scientific Study of Consciousness, Brussels, June 2000. [DED]
(2001) Inattentional awareness. *Psyche* 7(5). http://psyche.cs.monash.edu.au/v7/psyche-7-05-dulany.html [DED]
Dulany, D. E., Carlson, A. & Dewey, G. I. (1984) A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General* 113:541–55. [DED, aPP]
Duscherer, K. & Holender, D. (2002) No negative semantic priming from unconscious flanker words in sight. *Journal of Experimental Psychology: Human Perception and Performance* 28:839–53 [DH]
(submitted) The affective Simon effect: How automatic? How conscious? How specific? [DH]
Eich, E. (1984) Memory for unattended events: Remembering with and without awareness. *Memory and Cognition* 12:105–11. [aPP, GW]
Ellis, R. D. (1995) *Questioning consciousness: The interplay of imagery, cognition, and emotion in the human brain.* John Benjamins. [RDE]
Elman, J. L. (1990) Finding structure in time. *Cognitive Science* 14:179–211. [arPP]
Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plukett, K. (1996) *Rethinking innateness: A connectionist perspective on development.* MIT Press. [aPP]
Erdelyi, M. H. (1992) Psychodynamics and the unconscious. *American Psychologist* 47:784–87. [RAC]
Eriksen, C. W. (1995) The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition* 2:101–18. [DH]
Evans, J. St. B. T. (1989) *Bias in human reasoning: Causes and consequences.* Erlbaum. [PB, OH]
Evans, J. St. B., Newstead, S. E. & Byrne, P. (1993) *Human reasoning.* Erlbaum. [aPP]
Fahlman, S. E. (1991) The recurrent cascade-correlation architecture. In: *Advances in neural information processing systems* 3, ed. R. P. Lippmann, J. E. Moody & D. S. Touretzky. Morgan Kaufmann. [FD]
Fahlman, S. E. & Lebiere, C. (1990) The cascade-correlation learning architecture. In: *Advances in neural information processing systems, vol. 2,* ed. D. S. Touretzky. Morgan Kaufmann. [FD, aPP]
Farah, M. J. (1994) Visual perception and visual awareness after brain damage: A tutorial overview. In: *Attention and performance XV: Conscious and nonconscious information processing,* ed. C. Umilta & M. Moscovitch. MIT Press. [DED, aPP]
Field, H. (1978) Mental representation. *Erkenntnis* 13:9–61. [ZD]
Fischer, K. & Granott, N. (1995) Beyond one-dimensional change: Parallel, concurrent, socially distributed processes in learning and development. *Human Development* 38:302–14. [aPP]
Fodor, J. A. (1975) *The language of thought.* Crowell. [BS]
(1978) Propositional attitudes. *The Monist* 61:501–23. [ZD]
(1981) *Representations.* MIT Press. [RDE]
(1983) *The modularity of mind.* MIT Press. [aPP]
(1987) *Psychosemantics: The problem of meaning in the philosophy of mind.* MIT Press. [GO]
Fodor, J. A. & Pylyshyn, Z. W. (1981) How direct is visual perception?: Some reflections on Gibson's "Ecological Approach." *Cognition* 9:139–96. [GO]
Fong, G. T., Krantz, D. H. & Nisbett, R. E. (1986) The effects of statistical training on thinking about everyday problems. *Cognitive Psychology* 18:253–92. [aPP]
Frensch, P. A., Buchner, A. & Lin, J. (1994) Implicit learning of unique and ambiguous serial transitions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:567–84. [aPP]
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. [aPP]
Freud, S. (1915/1957) The unconscious. In: *The standard edition of the complete*


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

*psychological works of Sigmund Freud, vol. 14,* ed. and trans. J. Strachey. Hogarth Press. [DED]
Friedland, R. P. & Weinstein, E. A. (1977) Hemi-inattention and hemisphere specialization: Introduction and historical review. *Advance in Neurology* 18:1–31. [ES]
Frost, R. (1998) Toward a strong phonological theory of visual recognition: True issues and false trails. *Psychological Bulletin* 123:71–99. [JT]
Fulgosi, A. & Guilford, J. P. (1968) Short term incubation in divergent production. *American Journal of Psychology* 81:241–48. [aPP]
Gabrielli, J. D. E. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology* 49:87–115. [aPP]
Ganor-Stern, D., Seamon, J. G. & Carrasco, M. (1998) The role of attention and study time in explicit and implicit memory for unfamiliar visual stimuli. *Memory and Cognition* 26:1187–95. [aPP]
Garcia, J., Runsiniak, K. W. & Brett, L. P. (1977) Conditioning food-illness aversions in wild animals: Caveant canonici. In: *Operant-Pavlovian interactions,* ed. H. Davis & H. M. B. Hurwitz. Erlbaum. [aPP]
Gasser, M. & Smith, L. B. (1998) Learning nouns and adjectives: A connectionist account. *Language and Cognitive Processes* 13:269–306. [aPP]
Gibson, J. J. (1950) The concept of the stimulus in psychology. *American Psychologist* 15:694–705. [ZD]
(1960) The concept of the stimulus in psychology. *American Psychologist* 16:694–703. [BS]
(1966a) *The senses considered as a perceptual system.* Houghton-Mifflin. [rPP, BS]
(1966b) The problem of temporal order in stimulation and perception. *Journal of Psychology* 62:141–49. [BS]
(1968) What gives rise to the perception of motion? *Psychological Review* 75:335–46. [BS]
(1975) Events are perceivable but time is not. In: *The study of time II,* ed. J. T. Fraser & N. Lawrence. Springer Verlag. [BS]
(1979) *The ecological approach to visual perception.* Houghton-Mifflin. [BS]
Gick, M. L. & Holyoak, K. J. (1983) Schema induction and analogical transfer. *Cognitive Psychology* 15:1–38. [aPP]
Gillette, J., Gleitman, H., Gleitman, L. & Lederer, A. (1999) Human simulations of vocabulary learning. *Cognition* 73:135–76. [rPP]
Gombert, J. E. (1992) *Metalinguistic development.* University of Chicago Press. [CP]
Gomez, R. L. & Gerken, L. A. (1999) Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition* 70:109–35. [aPP]
Gomez, R. L., Gerken, L. A. & Schvaneveldt, R. W. (2000) The basis of transfer in artificial grammar learning. *Memory and Cognition* 28:253–63. [aPP]
Goodbody, S. J. & Wolpert, D. M. (1998) Temporal and amplitude generalization in motor learning. *Journal of Neurophysiology* 79:1825–938. [ASK]
Goodson, F. E. & Morgan, G. A. (1976) Evaluation of theory. In: *Theories in contemporary psychology,* ed. M. H. Marx & F. E. Goodson. Macmillan. [TY]
Green, R. E. A. & Shanks, D. R. (1993) On the existence of independent learning systems: An examination of some evidence. *Memory and Cognition* 21:304–17. [aPP]
Green, R. T. (1981) Beyond Turing. *Speculations in Science and Technology* 4(2):175–86. [rPP, MV]
Greenwald, A. G. (1992) New look 3: Unconscious cognition reclaimed. *American Psychologist* 47:766–79. [rPP, GW]
Greenwald, A. G., Draine, S. C. & Abrams, R. H. (1996) Three cognitive markers of unconscious semantic activation. *Science* 273:1699–702. [DH, aPP]
Grodzinsky, Y. (2000) The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences* 23:1–21. [aPP]
Groeger, J. A. (1984a) *Preconscious influences on language production.* Doctoral dissertation, Queen's University of Belfast, Ireland. [MV]
(1984b) Evidence of unconscious semantic processing from a forced error situation. *British Journal of Psychology* 75:305–14. [MV]
(1988) Qualitatively different effects of undetected and unidentified auditory primes. *Quarterly Journal of Experimental Psychology* 40A:323–29. [rPP, MV]
Gruber, R. P., Reed, D. R. & Block, J. D. (1968) Transfer of the conditioned GSR from drug to nondrug state without awareness. *Journal of Psychology* 70:149–55. [aPP]
Guzeldere, G. (1997) The many faces of consciousness: A field guide. In: *The nature of consciousness: Philosophical debates,* ed. N. Block, O. Flanagan & G. Guzeldere. MIT Press. [rPP]
Hahne, A. & Jescheniak, J. D. (2001) What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research* 11(2):199–212. [MR-G]
Haith, J. (1978) Visual competence in early infancy. In: *Handbook of sensory physiology, vol 3: Perception,* ed. R. Held, H. Leibowitz & H. L. Teuber. Springer-Verlag. [aPP]
Halford, G. S., Phillips, S. & Wilson, W. H. (submitted) Three levels of cognitive processes: Nonstructured, functionally structured, and symbolically structured. [SP]
Harlow, J. M. (1848) Passage of an iron rod through the head. *Boston Medical and Surgical Journal* 39:389–93. [OH]
Harnad, S. (1990) The symbol grounding problem. *Physica* D(42):335–46. [RM]
Harris, R. (1980) *The language makers.* Duckworth. [CP]
Hasher, L. & Zacks, R. T. (1979) Automatic and effortful processes in memory. *Journal of Experimental Psychology: General* 108:356–88. [aPP, JT]
Hebb, D. O. (1949) *The organization of behavior.* Wiley. [RMF]
Hinton, G. E. (1990) Mapping part-whole hierarchies into connectionist networks. *Artificial Intelligence* 46:47–75. [aPP, SP]
Holender, D. (1986) Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences* 9:1–23. [DH, AJL, arPP, MV]
(1992) Expectancy effects, congruity effects, and the interpretation of response latency measurement. In: *Analytic approaches to human cognition,* ed. J. Alegria, D. Holender, J. Junca de Morais & M. Radeau. North-Holland. [DH]
Holland, P. C. (1980) CS-US interval as a determinant of the form of the Pavlovian conditioned response. *Journal of Experimental Psychology: Animal Behavior Process* 6:155–74. [aPP]
Houdé, O. (2000) Inhibition and cognitive development: Object, number, categorization, and reasoning. *Cognitive Development* 15:63–73. [OH]
Houdé, O., Zago, L., Crivello, F., Moutier, S., Pineau, A., Mazoyer, B., et al. (2001) Access to deductive logic depends on a right ventromedial prefrontal area devoted to emotion and feeling: Evidence from a training paradigm. *NeuroImage* 14:1486–92. [OH]
Houdé, O., Zago, L., Mellet, E., Moutier, S., Pineau, A., Mazoyer, B., et al. (2000) Shifting from the perceptual brain to the logical brain: The neural impact of cognitive inhibition training. *Journal of Cognitive Neuroscience* 12:721–28. [OH, rPP]
Hsiao, A. T. & Reber, A. (1998) The role of attention in implicit sequence learning: Exploring the limits of the cognitive unconscious. In: *Handbook of implicit learning,* ed. M. Stadler & P. Frensch. Sage. [aPP]
Hubel, D. H. & Wiesel, T. N. (1959) Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology* 148:574–91. [RDE]
Jackendoff, R. (1987) *Consciousness and the computational mind.* MIT Press. [RDE]
Jackson, J. H. (1876/1932). *Selected writings of John Hughlings Jackson,* ed. James Taylor. Hodden & Stoughton. [ES]
Jacoby, L. L. & Kelley, C. M. (1987) Unconscious influences of memory. *Personality and Social Psychology Bulletin* 13:314–36. [rPP]
Jacoby, L .L., Ste-Marie, D. & Toth, J. P. (1993) Redefining automaticity: Unconscious influences, awareness and control. In: *Attention, selection, awareness and control. A tribute to Donald Broadbent,* A. D. Baddeley & L. Weiskrantz. Oxford University Press. [aPP]
Jacoby, L. L., Toth, J. P. & Yonelinas, A. P. (1993) Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General* 122:139–54. [GW]
Jacoby, L. L., Woloshyn, V. & Kelley, C. M. (1989) Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General* 118:115–25. [rPP]
Jacoby, L. L., Yonelinas, A. P. & Jennings, J. (1997) The relation between conscious and unconscious (automatic) influences. A declaration of independence. In: *Scientific approaches to consciousness,* ed. J. Cohen & J. W. Schooler. Erlbaum. [rPP]
Jakab, Z. (2000) Ineffability of qualia: A straightforward naturalistic explanation. *Consciousness and Cognition* 9:329–51. [rPP]
James, W. (1890/1950) *The principles of psychology.* Holt, Rinehart and Winston. [BS]
Jiang, Y. & Chun, M. M. (2001) Selective attention modulates implicit learning. *Quarterly Journal of Experimental Psychology* 54:1105–24. [aPP]
Jiménez, L. & Mendez, C. (1999) Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25:236–59. [aPP]
Johnson, E. K. & Jusczyk, P. W. (2001) Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language* 44:548–67. [CP]
Johnson-Laird, P. N. (1983) *Mental models.* Harvard University Press. [PB, aPP]
Johnson-Laird, P. N. & Byrne, R. M. J. (1991) *Deduction.* Erlbaum. [PB]
(2002) Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review* 109(4):646–78. [PB, rPP]
Jonides, J. (1981) Voluntary versus automatic control over the mind's eye's movement. In: *Attention and performance XI,* ed. J. Long & A. Baddeley. Erlbaum. [PBart]
Jusczyk, P. W. (1997) *The discovery of spoken language.* MIT Press. [aPP]
Jusczyk, P. W. & Aslin, R. N. (1995) Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology* 29:1–23. [aPP]



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Kahan, T. A. (2000) Negative priming from masked words: Retrospective prime clarification or center-surround inhibition? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26:1392–410. [DH]
Kahneman, D. & Treisman, A. (1984) Changing views of attention and automaticity. In: *Varieties of attention,* ed. R. Parasuraman & D. R. Davies. Academic Press. [aPP]
Kagan, J. (1971) *Change and continuity in infancy.* Wiley. [aPP]
Kandel, E. R. (1991) Cellular mechanisms of learning and the biological basis of individuality. In: *Principles of neuroscience,* ed. E. R. Kandel, J. H. Schwartz & T. M. Jessell. Elsevier. [GW]
Karmiloff-Smith, A. (1992) *Beyond modularity: A developmental perspective on cognitive science.* Bradford/ MIT Press. [aPP]
(1995) *Beyond modularity: A developmental perspective on cognitive science.* Bradford Books, MIT Press. [PB]
Ke, J.-Y., Minett, J. A., Au, C.-P. & Wang, W. S.-Y. (2002) Self-organization and selection in the emergence of vocabulary. *Complexity* 7(3):41–54. [WS-YW]
Kelso, J. A. S. (1990) Phase transitions: Foundations of behavior. In: *Synergetics of cognition,* ed. H. Haken & M. Stadler. Springer-Verlag. [rPP]
Kelso, J. A. S. & Schöner, G. S. (1988) Self-organization of coordinative movement patterns. *Human Movement Science* 7:27–46. [rPP]
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 . [rPP]
Kihlstrom, J. F. (1987) The cognitive unconscious. *Science* 237:1445–52. [aPP, JT, MV]
(1996) Perception without awareness of what is perceived, learning without awareness of what is learned. In: *The science of consciousness: Psychological, neuropsychological, and clinical reviews,* ed. M. Velmans. Routledge. [MV]
Klinger, M. R., Burton, P. C. & Pitts, G. S. (2000) Mechanisms of unconscious priming: I. Response competition, not spreading activation. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26:441–55. [DH]
Konorsky, J. (1967) *Integrative activity of the brain.* University of Chicago Press. [aPP]
Kosslyn, S. M. (1983) *Ghosts in the mind's machine: Creating and using images in the brain.* W. W. Norton. [RDE]
Kunst-Wilson, W. R. & Zajonc, R. B. (1980) Affective discrimination of stimuli that cannot be recognized. *Science* 207:557–58. [RFB]
Kurthen, M., Grunwald, T. & Elger, C. E. (1998) Will there be a neuroscientific theory of consciousness? *Trends in Cognitive Sciences* 2: 229–34. [MK]
LaBerge, D. & Samuels, S. J. (1974) Toward a theory of automatic information processing in reading. *Cognitive Psychology* 6:293–323. [aPP]
LeDoux, J. E. (1996) *The emotional brain.* Simon and Schuster. [GW]
Levine, A. & Bickhard, M. H. (1999) Concepts: Where Fodor went wrong. *Philosophical Psychology* 12:5–23. [RDE]
Levine, M. (1975) *A cognitive theory of learning.* Erlbaum. [CAS]
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. [aPP]
Lewicki, P., Hill, T. & Czyzewska, M. (1992) Nonconscious acquisition of information. *American Psychologist* 47:796–801. [aPP]
Libet, B., Gleason, C. A., Wright, E. W. & Pearl, D. K. (1983) Time of conscious intention to act in reaction to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain* 106:623–42. [rPP]
Lloyd, D. (1996) Consciousness, connectionism, and cognitive neuroscience: A meeting of the minds. *Philosophical Psychology* 9:61–79. [rPP]
Logan, G. D. (1988) Towards an instance theory of automatization. *Psychological Review* 76:165–78. [aPP, JT]
(1990) Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology* 22:1–35. [aPP]
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. [aPP]
Lovibond, P. F. & Shanks, D. R. (2002) The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes* 28:3–26. [aPP]
Ludden, D. & Gupta, P. (2000) Zen in the art of language acquisition: Statistical learning and the Less is More hypothesis. In: *Proceedings of the 22nd Annual Conference of the Cognitive Science Society,* ed. L. R. Gleitman & A. K. Joshi. Erlbaum. [aPP]
MacDonald, P. & MacLeod, C. M. (1998) The influence of attention at encoding on direct and indirect remembering. *Acta Psychologica* 98:298–310. [aPP]
Mackintosh, N. J. (1975) A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review* 82:276–98. [aPP]
(1997) Has the wheel turned full circle? Fifty years of learning theory, 1946–1996. *The Quarterly Journal of Experimental Psychology* 50A:879–98. [aPP]
MacLeod, C. M. (1991) Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin* 109:163–203. [DH]
MacWhinney, B. (1995) *The Childes project: Tolls for analysing talk.* Erlbaum. [aPP]
Mandler, G. (1975) Consciousness: Respectable, useful, and probably necessary. In: *Information processing and cognition: The Loyola symposium,* ed. R. Solso. Erlbaum. [aPP]
(1994) Hyperamnesia, incubation and mind popping: On remembering without really trying. In: *Attention and performance XV: Conscious and nonconscious information processing,* ed. C. Umilta & M. Moscovitch. MIT Press. [aPP]
Mandler, G., Nakamura, Y. & Van Zandt, B. T. S. (1987) Nonspecific effects of exposure on stimuli that cannot be recognized. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:646–48. [rPP]
Manza, L. & Reber. A. S. (1997) Representing artificial grammars: Transfer across stimulus forms and modalities. In: *How implicit is implicit learning,* ed. D. C. Berry. Oxford University Press. [LL, aPP]
Manzotti, R. & Sandini, G. (2001) Does functionalism really deal with the phenomenal side of experience? *Behavioral Brain Science* 24(5). [RM]
Manzotti, R. & Tagliasco, V. (2001) *Coscienza e realtà. Una teoria della coscienza per costruttori e studiosi di menti e cervelli.* Il Mulino. [RM]
Marcel, A. J. (1983) Conscious and unconscious perception: Experiments in visual masking and word recognition. *Cognitive Psychology* 15:197–237. [aPP]
(1986) Consciousness and processing: Choosing and testing a null hypothesis. *Behavioral and Brain Sciences* 9:40–41. [aPP]
(1988) Phenomenal experience and functionalism. In: *Consciousness in contemporary science,* ed. A. J. Marcel & E. Bisiach. Oxford University Press. [PBart]
Marcus, G. F. (1998) Rethinking eliminative connectionism. *Cognitive Psychology* 37:243–82. [SP]
(1999) Connectionism: With or without rules? *Trends in Cognitive Science* 3:168–70. [TY]
Marcus, G. F., Vijayan, S., Rao, S. B. & Vishton, P. M. (1999) Rule learning by seven-month-old infants. *Science* 283:77–80. [FD, aPP, SP, TY]
Marescaux, P.-J., Dejean, K. & Karnas, G. (1990) *Acquisition of specific or general knowledge at the control of a dynamic simulated system: An evaluation through a static situations questionnaire and a transfer control task.* Report 2PR2GK of the KAUDYTE project (ESPRIT BRA #3219). [aPP]
Markman, E. M. (1990) Constraints children place on word meanings. *Cognitive Science* 14:57–77. [arPP]
Markovits, H. & Barrouillet, P. (2002a) The development of conditional reasoning: A mental model account. *Developmental Review* 22:5–36. [PB, rPP]
(2002b) The development of conditional reasoning with familiar premises: A computational model. (submitted). [PB]
Marr, D. (1982) *Vision: A computational investigation into the human representation and processing of visual information.* W. H. Freeman. [AJL]
Marx, M. H. (1976) Formal theory. In: *Theories in contemporary psychology,* ed. M. H. Marx & F. E. Goodson. Macmillan. [TY]
Mathews, R. C., Buss, R. R., Stanley, W. B., Blancard-Fields, F., Cho, J.-R. & Druhan, B. (1989) Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:1083–100. [aPP]
Mathis, D. & Mozer, M. (1996) Conscious and unconscious perception: A computational theory. In: *Proceedings of the 18th Annual Conference of the Cognitive Science Society,* ed. G. Cottrell. LEA. [RMF]
McClelland, J. L. & Plaut, D. C. (1999) Does generalization in infant learning implicate abstract algebra-like rules? *Trends in Cognitive Science* 3:166–68. [aPP, TY]
McCloskey, M. (1992) Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition* 44:107–57. [aPP]
McDonald, J. L. (1997) Language acquisition: The acquisition of linguistic structure in normal and special populations. *Annual Review of Psychology* 48:215–41. [aPP]
McDonald, S. & Ramscar, M. (2001) Testing the distributional hypothesis: The influence of context on judgments of semantic similarity. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society.* University of Edinburgh. [aPP]
McGeorge, P. & Burton, A. M. (1990) Semantic processing in an incidental learning task. *Quarterly Journal of Experimental Psychology* 42A:597–609. [aPP]
McGlinchey-Berroth, R., Milberg, W. P., Verfaellie, M., Grande, L., D'Esposito, M. & Alexander, M. (1996) Semantic processing and orthographic specificity in hemispatial neglect. *Journal of Cognitive Neuroscience* 8:291–304. [ES]
McHugh, D. E. & Bahill, A. T. (1985) Learning to track predictable target waveforms without a time delay. *Investigative Ophthalmology and Visual Science* 26:932–37. [RDE]
McKoon, G. & Ratcliff, R. (1998) Memory-based language processing: Psycholinguistic research in the 1960s. *Annual Review of Psychology* 49:25–32. [aPP]
McLeod, P. & Dienes, Z. (1993) Running to catch the ball. *Nature* 362: 23. [ASK, aPP]


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

(1996) Do fielders know where to go to catch the ball or only how to get there? *Journal of Experimental Psychology: Human Perception and Performance* 22:531–43. [ASK]
Merikle, P. M. (1992) Perception without awareness. *American Psychologist* 47:792–95. [GW]
Merikle, P. M. & Joordens, S. (1997) Parallels between perception without attention and perception without awareness. *Consciousness and Cognition* 6:219–36. [aPP, GW]
Merikle, P. M. & Reingold, E. M. (1998) On demonstrating unconscious perception: Comment on Draine and Greenwald (1998). *Journal of Experimental Psychology: General* 127: 304–10. [aPP]
Merikle, P. M., Smilek, D. & Eastwood, J. D. (2001) Perception without awareness: Perspectives from cognitive psychology. *Cognition* 79(1–2):115–34. [PBart]
Merleau-Ponty, M. (1942) *La structure du comportement.* Presses Universitaires de France. [PBart]
(1962) *The phenomenology of perception.* Routledge and Kegan Paul. [rPP, BS]
Michaels, C. F. & Oudejans, R. D. (1992) The optics and actions of catching fly balls: Zeroing out optic acceleration. *Ecological Psychology* 4:199–222. [ASK]
Miller, G. A. (1958) Free recall of redundant strings of letters. *Journal of experimental Psychology* 56:485–91. [aPP]
(1962) *Psychology: The science of mental life.* Harper and Row. [aPP]
Miller, G. H. (1956) The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63:81–97. [BS]
Miller, J. (2000) Measurement error in subliminal perception experiments: Simulation analyses of two regression models. *Journal of Experimental Psychology: Human Perception and Performance* 26:1461–77. [aPP]
Miller, J. & Weinert, R. (1998) *Spontaneous spoken language.* Clarendon Press. [CP]
Millikan, R. G. (1993) *White Queen psychology and other essays for Alice.* Bradford Books/MIT Press. [ZD]
Mulligan, N. W. & Hirshman, E. (1997) Measuring the bases of recognition memory: An investigation of the process-dissociation framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:280–304. [rPP]
Munakata, Y. (2001) Graded representations in behavioral dissociations. *Trends in Cognitive Sciences* 5:309–15. [rPP]
Murphy, S. T. & Zajonc, R. B. (1993) Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology* 64:723–39. [GW]
Murray, D. J. (2001). Partial matching theory and the memory span. *Behavioral and Brain Sciences* 24:133–34. [DJM]
Murray, D. J., Boudreau, N., Burggraf, K. K., Dobell, L., Guger, S. L., Leask, A., Stanford, L., Tate, T. L. & Wheeler, M. (1999) A grouping interpretation of the modality effect in immediate probed recognition. *Memory and Cognition* 27:234–45. [DJM]
Murray, D. J., Burhop, J., Centa, S., Chande, N., Oinonen, K., Thomas, T., Wilkie, T. & Farahmand, B. (1998) A partial matching theory of the mirror effect in immediate probed recognition. *Memory and Cognition* 26:1196–213. [DJM]
Naccache, L. & Dehaene, S. (2001) Unconscious semantic priming extends to novel unseen stimuli. *Cognition* 80:223–37. [DH]
Neal, A. & Hesketh, B. (1997) Episodic knowledge and implicit learning. *Psychonomic Bulletin and Review* 4:24–37. [aPP]
Neely, J. H. (1991) Semantic priming effects in visual word recognition: A selective review of current findings and theories. In: *Basic processes in reading: Visual word recognition,* ed. D. Besner & G. W. Humphreys. Erlbaum. [DH]
Neely, J. H. & Kahan, T. A. (2001) Is semantic activation automatic? A critical re-evaluation. In: *The nature of remembering: Essays in honor of Robert G. Crowder,* ed. H. L. Roediger & J. S. Nairne. Science Conference Series. American Psychological Association. [aPP]
Neumann, O. (1984) Automatic processing: A review of recent findings and a plea for an old theory. In: *Cognition and motor processes,* ed. W. Prinz & A. F. Sanders. Springer. [aPP]
Newell, A., Shaw, J. C. & Simon, H. A. (1960) Report on a general problem solving program for a computer. In: *Information processing: Proceedings of the International Conference on Information Processing.* UNESCO. [MV]
Newell, K. M. & McDonald, P. V. (1992) Searching for solutions to the coordination function: Learning as exploratory behavior. In: *Tutorials in motor behavior II,* ed. G. E. Stelmach & J. Requin. Elsevier/North Holland. [rPP]
Newport, E. (1990) Maturational constraints on language learning. *Cognitive Science* 14:11–28. [aPP]
Newton, N. (2002) Representation in theories of embodied cognition. *Theoria* (in press). [RDE]
Nissen, M. J. & Bullemer, P. (1987) Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology* 19:1–32. [aPP]
Noë, A., Pessoa, L. & Thompson, E. (2000) Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition* 7:93–106. [aPP]
Nowak, M. A., Komarova, N. L. & Niyogi, P. (2001) Evolution of universal grammar. *Science* 291:114–18. [WS-YW]
Oaksford, M. & Chater, N. (2001) The probabilistic approach to human reasoning. *Trends in Cognitive Sciences* 5(8):349–57. [PB, rPP]
O'Brien, G. & Opie, J. (1999a) A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences* 22:127–48. [RAC, AD, GO, arPP]
(1999b) Putting content into a vehicle theory of consciousness. *Behavioral and Brain Sciences* 22:175–96. [AD, aPP]
(1997) Cognitive science and phenomenal consciousness. *Philosophical Psychology* 10:269–86. [RM]
(1999) A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences* 21:127–96. [RM]
Ong, W. J. (1982) *Orality and literacy : The technologizing of the word.* Methuen. [CP]
O'Regan, J. K. (1992) Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology* 46:461–88. [CP, aPP]
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: 401–26. [aPP]
Paivio, A. (1986) *Mental representations: A dual coding approach.* Oxford University Press. [RDE]
Palmer, S. E. (1999) *Vision science: Photons to phenomenology.* MIT Press. [GO]
Parkin, A. J. & Russo, R. (1990) Implicit and explicit memory and the automatic/effortful distinction. *European Journal of Psychology* 2:71–80. [aPP]
Perner, J. (1991) *Understanding the representational mind.* MIT Press. [ZD]
1998) The meta-intentional nature of executive functions and theory of mind. In: *Language and thought: Interdisciplinary themes,* ed. P. Carruthers & J. Boucher. Cambridge University Press. [ZD]
(in press) Dual control and the causal theory of action. In: *Agency and self-awareness,* ed. N. Eilan & J. Roessler. Oxford University Press. [ZD]
Perner, J. & Dienes, Z. (in press). Developmental aspects of consciousness: How much theory of mind do you need to be consciously aware? *Consciousness and Cognition.* [ZD]
Perruchet, P. (1984) Dual nature of anticipatory classically conditioned reactions. In: *Preparatory states and processes,* ed. S. Kornblum & J. Requin. Erlbaum. [aPP]
(1994) Learning from complex rule-governed environments: On the proper functions of nonconscious and conscious processes. In: *Attention and performance XV: Conscious and nonconscious information processing,* ed. C. Umilta & M. Moscovitch. MIT Press.[aPP]
Perruchet, P. & Amorim, M. A. (1992) Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:785–800. [aPP]
Perruchet, P. & Gallego, J. (1997) A subjective unit formation account of implicit learning. In: *How implicit is implicit learning,* ed. D. Berry. Oxford University Press. [LJ, aPP]
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. [aPP]
Perruchet, P. & Pacteau, C. (1990) Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General* 119: 264–75. [arPP]
Perruchet, P., Pacteau, C. & Gallego, J. (1997a) Abstraction of covariation in incidental learning and covariation bias. *British Journal of Psychology* 88:441–58. [aPP]
Perruchet, P. & Vinter, A. (1998a) Learning and development: The implicit knowledge assumption reconsidered . In: *Handbook of implicit learning,* ed. M. Stadler & P. Frensch. Sage. [aPP]
(1998b) PARSER: A model for word segmentation. *Journal of Memory and Language* 39:246–63. [CP, arPP]
Perruchet, P., Vinter, A. & Gallego, J. (1997b) Implicit learning shapes new conscious percepts and representations. *Psychonomic Bulletin and Review* 4:43–48. [arPP]
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. [aPP]
Peters, A. M. (1983) *The units of language acquisition.* Cambridge University Press. [CP]
Phaf, R. H. (2001) Only explicit memory seems affected by depth of hypnotic state. *Anesthesiology* 94:380–81. [GW]
Phaf, R. H. & Wolters, G. (1997) A constructivist and connectionist view on conscious and nonconscious processes. *Philosophical Psychology* 10:287–307. [GW]



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Phillips, S. (1999) Systematic minds, unsystematic models: Learning transfer in humans and networks. *Minds and Machines* 9(3):383–98. [SP]
(2000) Constituent similarity and systematicity: The limits of first-order connectionism. *Connection Science* 12(1):45–63. [SP]
Piaget, J. (1936/1953) *The origin of intelligence in the child.* Routledge and Kegan Paul. [ZD]
(1984) Piaget's theory. In: *Handbook of child psychology: History, theories, and methods, vol. 1, 4th edition,* ed. P. H. Mussen & W. Kessen. Wiley. [OH]
(1985) *The equilibration of cognitive structures: The central problem of intellectual development.* University of Chicago Press. [aPP]
Pinker, S. (1984) *Language learnability and language development.* Harvard University Press. [MR]
(1991) Rules of language. *Science* 253:530–35. [rPP]
(1999) Out of the minds of babes. *Science* 283:40–41. [aPP]
Poggio, T. & Koch, C. (1985) Ill-posed problems in early vision from computational theory to analogue networks. *Proceedings of the Royal Society of London* B226:303–23. [RDE]
Posner, M. I. & Boies, S. J. (1971) Components of attention. *Psychological Review* 78:391–408. [aPP]
Posner, M. I. & Cohen, Y. (1984) Components of visual orienting. In: *Attention and performance X,* ed. H. Bouma & D. G. Bouwhis. Erlbaum. [JT]
Posner, M. I. & Snyder, C. R. R. (1975) Attention and cognitive control. In: *Information processing and cognition: The Loyola symposium,* ed. R. Solso. Erlbaum. [PBart]
Posner, M. I., Walker, J. A., Friedrich, F. J. & Rafal, R. D. (1984) Effects of parietal injury on covert orienting of attention. *The Journal of Neuroscience* 4:1863–74. [ES]
Pribam, K. H. (1971) *Languages of the brain.* Prentice Hall. [RDE]
(1980) Mind, brain, and consciousness. In: *The psychobiology of consciousness,* ed. J. M. Davidson & R. J. Davidson. Plenum Press. [aPP]
(1991) *Brain and perception: Holonomy and structure in figural processing.* Erlbaum. [RDE]
Prinz, W. (1997) Why Donders has led us astray. In: *Theoretical issues in stimulus-response compatibility,* ed. B. Hommel & W. Prinz. North-Holland. [DH]
Pylyshyn, Z. (1973) What the mind's eye tells the mind's brain. *Psychological Bulletin* 80:1–23. [RDE]
Ramscar, M. (2002) The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology* 45:45–94. [rPP]
Ramsey, W. (1997) Do connectionist representations earn their explanatory keep? *Mind and Language* 12:34–66. [rPP]
Reber, A. S. (1967) Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6:855–63. [aPP]
(1969) Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology* 81:115–19. [aPP]
(1976) Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory* 2:88–94. [aPP]
(1989) Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General* 118:219–35. [rPP]
(1992a) An evolutionary context for the cognitive unconscious. *Philosophical Psychology* 5:33–51. [LL]
(1992b) The cognitive unconscious: An evolutionary perspective. *Consciousness and Cognition* 1:93–133. [LL]
(1993) *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* Oxford University Press. [LL, aPP]
Reber, A. S. & Allen, R. (1978) Analogy and abstraction strategies in synthetic grammar learning: A functionalist interpretation. *Cognition* 6:189–221. [LL]
(2000) Individual differences in implicit learning. In: *Individual differences in conscious experience,* ed. R. G. Kunzendorf & B. Wallace. John Benjamins. [LL]
Reber, A. S. & Reber, E. S. (2001) *Dictionary of psychology, 3rd edition.* Penguin/Viking. [LL]
Redington, M. & Chater, N. (2002) Knowledge representation and transfer in artificial grammar learning. In: *Implicit learning and consciousness: An empirical, philosophical, and computational consensus in the making,* ed. R. M. French & A. Cleeremans. Psychology Press. [aPP]
Redington, M., Chater, N. & Finch, S. (1998) Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science* 22:425–69. [aPP, MR]
Reed, J. M. & Johnson, P. J. (1998) Implicit learning: Methodological issues and evidence of unique characteristics. In: *Handbook of implicit learning,* ed. M. Stadler & P. Frensch. Sage. [rPP]
Reeves, L. M. & Weisberg, R. W. (1990) Analogical transfer in problem solving: Schematic representations and cognitive processes. Paper presented at the meeting of the Eastern Psychological Association, Philadelphia, PA. [aPP]
(1994) The role of content and abstract information in analogical transfer. *Psychological Bulletin* 115:381–400. [aPP]
Reingold, E. & Merikle, P. M. (1998) Using direct and indirect measures to study perception without awareness. *Perception and Psychophysics* 44:563–75. [DED, GW]
Reingold, E. M. & Toth, J. P. (1996) Process dissociation versus task dissociations: A controversy in progress. In: *Implicit cognition,* ed. G. Underwood. Oxford University Press. [aPP]
Rey, G. (1991) Reasons for doubting the existence of even epiphenomenal consciousness. *Behavioral and Brain Sciences* 14:691–92. [aPP]
Richardson, J. (1991) Imagery and the brain. In: *Imagery and cognition,* ed. Cesare Cornoldi & Mark McDaniel. Springer-Verlag. [RDE]
Richardson-Klavehn, A. & Bjork, R. A. (1988) Measures of memory. *Annual Review of Psychology* 39:475–543. [GW]
Richardson-Klavehn, A., Gardiner, J. M. & Java, R. I. (1996) Memory: Task dissociation, process dissociations, and dissociations of consciousness. In: *Implicit cognition,* ed. G. Underwood. Oxford University Press. [aPP]
Rips, L. J. (1994) *The psychology of proof: Deductive reasoning in human thinking.* MIT Press. [PB]
Rivest, F. & Shultz, T. R. (2002) Application of knowledge-based cascade-correlation to vowel recognition. *IEEE International Joint Conference on Neural Networks.* [FD]
Rock, I. (1984) *Perception.* W. H. Freeman. [rPP]
Roediger, H. L. & McDermott, K. B. (1993) Implicit memory in normal human subjects. In: *Handbook of neuropsychology, vol. 8,* ed. F. Boller & I. Grafman. Elsevier. [aPP, GW]
Rosenthal, D. M. (1986) Two concepts of consciousness. *Philosophical Studies* 49:329–59. [ZD]
(1993) Thinking that one thinks. In: *Consciousness: Psychological and philosophical essays,* ed. M. Davies & G. W. Humphreys. Blackwell. [RAC]
(2000) Consciousness, content, and metacognitive judgments. *Consciousness and Cognition* 9:203–14. [ZD]
Ross, B. H. & Kennedy, P. T. (1990) Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Human Learning and Memory* 16:42–55. [aPP]
Ross, J. R. (1979) Where's English? In: *Individual differences in language ability and language behavior,* ed. C. J. Fillmore, D. Kempler & W.S-Y. Wang. Academic Press. [WS-YW]
Rotteveel, M., de Groot, P., Geutskens, A. & Phaf, R. H. (2001) Stronger suboptimal than optimal affective priming? *Emotion* 1:348–64. [GW]
Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996a) Statistical learning by 8-month-old infants. *Science* 274:1926–28. [CP, aPP]
Saffran, J. R., Johnson, E. K., Aslin, R. N. & Newport, E. L. (1999) Statistical learning of tone sequences by human infants and adults. *Cognition* 70:27–52. [aPP]
Saffran, J. R., Newport, E. L. & Aslin, R. N. (1996b) Word segmentation: The role of distributional cues. *Journal of Memory and Language* 35:606–21. [aPP]
Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A. & Barrueco, S. (1997) Incidental language learning. *Psychological Science* 8:101–105. [CP, aPP]
Schmidt, P. A. & Dark, V. J. (1998) Attentional processing of "unattended" flankers: Evidence for a failure of selective attention. *Perception and Psychophysics* 60:227–38. [aPP]
Schneider, W. & Shiffrin, R. (1977) Controlled and automatic human information processing: I. Detection search and attention. *Psychological Review* 84:127–90. [JT]
Schoenemann, P. T. & Wang, W. S.-Y. (1996) Evolutionary principles and the emergence of syntax. *Behavioral and Brain Sciences* 19:646–47. [WS-YW]
Schooler, J. W. (2002) Re-presenting consciousness: Dissociations between experience and meta-consciousness. *Trends in Cognitive Sciences* 6:339–44. [rPP]
Schweinberger, S. R. & Stief, V. (2001) Implicit perception in patients with visual neglect: Lexical specificity in repetition priming. *Neuropsychologia* 39:420–29. [ES]
Schyns, P. G., Golstone, R. L & Thibaut, J.-P. (1998) The development of features in object concepts. *Behavioral and Brain Sciences* 21:1–53. [aPP]
Seamon, J. G., Brody, N. & Kauff, D. M. (1983) Affective discrimination of stimuli that are not recognized: Effect of delay between study and test. *Bulletin of the Psychonomic Society* 21:187–89. [RFB]
Searle, J. R. (1983) *Intentionality, an essay in the philosophy of mind.* Cambridge University Press. [RM]
(1990) Consciousness, explanatory inversion, and cognitive science. *Behavioral and Brain Sciences* 13:585–642. [DH, rPP]
(1992) *The rediscovery of the mind.* MIT Press. [AD, DH, rPP]
(1999) Chinese room argument. In: *The MIT encyclopedia of the cognitive sciences,* ed. R. A. Wilson & F. C. Kail. MIT Press. [LJ]
Seger, C. A. (1997) Multiple forms of learning in implicit learning. In: *Handbook of implicit learning,* ed. M. Stadler & P. Frensch. Sage. [CAS]
Seidenberg, M. S. & Elman, J. L. (1999) Do infants learn grammar with algebra or statistics? *Science* 284:433. [aPP]
Seligman, M. E. P. (1970) On the generality of the laws of learning. *Psychological Review* 77:406–18. [aPP]


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

Servan-Schreiber, D. & Anderson, J. R. (1990) Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16:592–608. [aPP, MR]
Shadmehr, R. & Mussa-Ivaldi, F. A. (1994) Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience* 14(5):3208–24. [ASK]
Shanks, D. (1995) *The psychology of associative learning.* Cambridge University Press. [aPP]
Shanks, D. R. & Channon, S. (in press) Effects of a secondary task on "implicit" sequence learning: Learning or performance? *Psychological Research.* [aPP]
Shanks, D. R., Johnstone, T. & Staggs, L. (1997) Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology,* 50A:216–52. [aPP]
Shanks, D. R. & St. John, M. F. (1994) Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences* 17:367–47. [arPP, MR]
Shanon, B. (1993) The representational and the presentational: An essay on cognition and the study of mind. Harvester-Wheatsheaf. [BS]
Shastri, L. & Ajjanagadde, V. (1993) From simple associations to systematic reasoning. *Behavioral and Brain Sciences* 16:417–94. [sPP]
Shaw, R. & Pittenger, J. (1978) Perceiving change. In: *Models of perceiving and processing information,* ed. R. Pick & E. Saltzman. Erlbaum. [BS]
Shaw, R. & Todd, J. (1980) Abstract machine theory and direct perception. *Behavioral and Brain Sciences* 3:400–401. [BS]
Shevrin, H. & Dickman, S. (1980) The psychological unconscious. *American Psychologist* 35:421–34. [aPP]
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. [aPP]
Shultz, T. R. (1994) The challenge of representational redescription. *Behavioral and Brain Sciences* 17(4):728–29. [FD]
(in press) *Computational developmental psychology.* MIT Press. [FD]
Shultz, T. R. & Bale, A. C. (2001) Neural network simulation of infant familiarization to artificial sentences: Rule-like behavior without explicit rules and variables. *Infancy* 2:501–36. [FD]
Shultz, T. R. & Rivest, F. (2001) Knowledge-based cascade-correlation: Using knowledge to speed learning. *Connection Science* 13:43–72. [FD]
Siéroff, E. (1991) Perception of visual letter strings in a case of left neglect : Manipulation of the word form. *Brain and Language* 41:565–89. [ES]
Siéroff, E. & Michel, F. (1987) Verbal visual extinction in right/left hemisphere lesion patients and the problem of lexical access. *Neuropsychologia* 25:907–18. [ES]
Siéroff, E., Pollatsek, A. & Posner, M. I. (1988) Recognition of visual letter strings following injury to the posterior visual spatial attention system. *Cognitive Neuropsychology* 5:427–49. [ES]
Silva-Pereyra, J., Rivera-Gaxiola, M., Klarman, L. & Kuhl, P. K. (2003) Semantic and grammatical processing in toddlers: Evidence from event-related brain potentials. Poster to be presented at the Cognitive Neuroscience Society Meeting, New York, USA, March 2003. [MR-G]
Singley, M. K. & Anderson, J. R. (1989) *Transfer of cognitive skill.* Harvard University Press. [aPP]
Sloman, S. A. (1996) The empirical case for two systems of reasoning. *Psychological Bulletin* 119:3–22. [aPP, RS]
Smith, E. E., Langston, C. & Nisbett, R. E. (1992) The case for rules in reasoning. *Cognitive Science* 16:1–40. [aPP]
Smith, E. E., Patalano, A. L. & Jonides, J. (1998) Alternative strategies of categorization. *Cognition* 65:167–96. [CAS]
Smith, S. M. & Blankenship, S. E. (1989) Incubation effects. *Bulletin of the Psycholomic Society* 27:311–14. [aPP]
(1991) Incubation and the persistence of fixation in problem solving. *American Journal of Psychology* 104:61–87. [aPP]
Smolensky, P. (1988) On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11:1–74. [aPP, RS]
Spelke, E. S., Breinlinger, K., Macomber, J. & Jacobson, K. (1992) Origins of knowledge. *Psychological Review* 99:605–32. [aPP]
Spence, K. W. (1937) The differential response in animals to stimuli varying within a single dimension. *Psychological Review* 44:430–44. [aPP]
Sperling, G. (1960) The information available in brief visual presentations. *Psychological Monographs* 74(11):1–29. [PBart]
Sperry, R. (1965) Mind, brain, and humanist values. In: *New views of the nature of man,* ed. J. R. Platt. University of Chicago Press. [aPP]
(1991) In defense of mentalism and emergent interaction. *Journal of Mind and Behavior* 12:221–45. [arPP]
Squire, L. R. & Frambach, M. (1990) Cognitive skill learning in amnesia. *Psychobiology* 18: 109–17. [aPP]
Srebro, R. (1985) Localization of visually evoked cortical activity in humans. *Journal of Physiology* 360:233–46. [RDE]
Stadler, M. A. (1989) On learning complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:1061–69. [aPP]
(1992) Statistical structure and implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:318–27. [aPP]
(1995) Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21:674–85. [aPP]
Stadler, M. A. & Frensch, P. A., eds. (1998) *Handbook of implicit learning.* Sage. [aPP]
Stadler, M. A., Warren, J. L. & Lesch, S. L. (2000) Is there cross-format transfer in implicit invariance learning? *The Quarterly Journal of Experimental Psychology* 53A:235–45. [aPP]
Steels, L. (1997) The synthetic modeling of language origins. *Evolution of Communication* 1:1–34. [WS-YW]
Sternberg, S. (1966) High-speed scanning in human memory. *Science* 153:652–54. [DJM]
Sun, R. (1996) Commonsense reasoning with rules, cases, and connectionist models: A paradigmatic comparison. *Fuzzy Sets and Systems* 82:187–200. [RS]
(1999) Accounting for the computational basis of consciousness: A connectionist approach. *Consciousness and Cognition* 8:529–65. [RS]
(2000) Symbol grounding: A new look at an old idea. *Philosophical Psychology* 13(2):149–72. [RS]
(2001) Computation, reduction, and teleology of consciousness. *Cognitive Systems Research* 1(4):241–49. [RS]
(2002) *Duality of the mind.* Erlbaum. [RS]
Sun, R., Merrill, E. & Peterson, T. (2001) From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science* 25(2):203–44. [RS]
Thelen, E., Schoner, G. & Scheier, C. (2001) The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences* 24(1):1–86. [RDE]
Thelen, E. & Smith, L. B. (1994) *A dynamic systems approach to the development of cognition and action.* MIT Press. [rPP]
Thivierge, J.-P. & Shultz, T. R. (2002) Finding relevant knowledge: KBCC applied to splice-junction determination. *IEEE International Joint Conference on Neural Networks.* [FD]
Thompson, E., ed. (2001). *Between ourselves. Second-person issues in the study of consciousness.* Imprint Academic. [RM, rPP]
Thompson, E. & Varela, F. J. (2001) Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences* 5:418–25. [arPP]
Tolman, E. C. (1932) *Purposive behaviors in animals and man.* Appleton-Century. [aPP]
Tomasello, M. (2000a) Do young children have adult syntactic competence? *Cognition* 74:209–53. [CP]
(2000b) The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences* 4:156–64. [CP, aPP]
Tooby, J. & Cosmides, L. (2000) Toward mapping the evolved functional organization of mind and brain. In: *The new cognitive neurosciences, 2nd. edition,* ed. M. Gazzaniga. MIT Press. [OH]
Treisman, A. M. (1998) Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society* B353:1295–1306. [ES]
Treisman, A. M. & Gelade, G. (1980) A feature integration theory of attention. *Cognitive Psychology* 12:97–136. [aPP]
Tunney, R. J. & Altmann, G. T. M. (1999) The transfer effect in artificial grammar learning: Re-appraising the evidence of transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25:1322–33. [aPP]
Turvey, M. T., Shaw, R. E., Reed, E. S. & Mace, W. M. (1981) Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn. *Cognition* 3:237–304. [BS]
Tzelgov, J. (1997a) Automatic but conscious: That is how we act most of the time. In: *The automaticity of everyday life: Advances in social cognition, vol. X,* ed. R. S. Wyer. Erlbaum. [aPP, JT]
(1997b) Specifying the relations between automaticity and consciousness: A theoretical note. *Consciousness and Cognition* 6:441–51. [JT]
Tzelgov, J., Henik, A., Sneg, R. & Baruch, O. (1996) Unintentional reading via the phonological route: The Stroop effect with cross-script homophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:336–49. [JT]
Tzelgov, J., Porat, Z. & Henik, A. (1997) Automaticity and consciousness: Is perceiving the word necessary for reading it? *American Journal of Psychology* 110:429–48. [aPP]
Tzelgov, J., Yehene, V., Kotler, L. & Alon, A. (2000) Automatic comparisons of artificial digits never compared: Learning linear order relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26:106–20. [JT]
Varela, F. (2000) *Neurophenomenology.* Imprint Academic. [RM, rPP]
Velmans, M. (1991) Is human information processing conscious? *Behavioral and Brain Sciences* 14:651–726. [aPP, MV]



https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press

(1998) Goodbye to reductionism. In: *Toward a science of consciousness II,* ed. S. R. Hameroff, A. W. Kaszniak & A. C. Scott. MIT Press. [aPP]
(1999) When perception becomes conscious. *British Journal of Psychology* 90:543–66. [aPP]
(2000) *Understanding consciousness.* Routledge/Psychology Press. [MV]
(2002a) How could conscious experiences affect brains? (target article) *Journal of Consciousness Studies: Special Issue on How Could Experiences Affect Brains?* 9(11): 3–29. [MV]
(2002b) Making sense of causal interactions between consciousness and brain (reply to commentaries). *Journal of Consciousness Studies: Special Issue on How Could Experiences Affect Brains?* 9(11): 69–95. [MV]
Venet, M. & Markovits, H. (2001) Understanding uncertainty with abstract conditional premises. *Merrill Palmer Quarterly* 47(1):74–99. [PB, rPP]
Vermersch, P. (2000) *Conscience directe et conscience réfléchie.* Manuscript available in: http://www.grex-fr.net/. [PBart]
Vinter, A. (1986) The role of movement in eliciting early imitations. *Child Development* 57:66–71. [aPP]
Vinter, A. & Perruchet, P. (1994) Is there an implicit level of representation? *Behavioral and Brain Sciences* 17:730–31. [rPP]
(1999a) Isolating unconscious influences: The neutral parameter procedure. *Quarterly Journal of Experimental Psychology* 52A:857–76. [ZD]
(1999b) What about consciousness during learning? *Behavioral and Brain Sciences* 22:176. [rPP]
Volpe, B. T., Ledoux, J. E. & Gazzaniga, M. S. (1979) Information processing of visual stimuli in an "extinguished" field. *Nature* 282:722–24. [ES]
Wagner, A. R. (1981) SOP: A model of automatic memory processing in animal behavior. In: *Information processing in animals: Memory mechanisms,* ed. N. E. Spear & R. R. Miller. Erlbaum. [aPP]
Wallas, G. (1926) *The art of thought.* Harcourt, Brace. [aPP]
Wang, W. S.-Y. (1991) Language prefabs and habitual thought. In: *Explorations in language,* W. S.-Y. Wang. Pyramid Press. [WS-YW]
Weiskrantz, L. (1986) *Blindsight: A case study and implications.* Oxford University Press. [RDE]
(1997) *Consciousness lost and found: A neuropsychological exploration.* Oxford University Press. [DED, AJL, arPP]
Whittlesea, B. W. A. & Dorken, M. D. (1993) Incidentally, things in general are incidentally determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General* 122:227–48. [LL, aPP]
Whittlesea, B. W. A. & Price, J. R. (2001) Implicit/explicit memory versus analytic/nonanalytic processing: Rethinking the mere exposure effect. *Memory and Cognition* 29:234–46. [DH, rPP]
Whittlesea, B. W. A. & Wright, R. L. (1997) Implicit (and explicit) learning: Acting adaptively without knowing the consequences. *Journal of Experimental Psychology: Learning, Memory and Cognition* 23:181–200. [LL, aPP]
Willingham, D. B. (1999) The neural basis of motor-skill learning. *Current Directions in Psychological Science* 8(6):178–82. [ASK]
Willingham, D. B., Nissen, M. J. & Bullemer, P. (1989) On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition* 15: 1047–60. [aPP]
Willingham, D. B., Wells, L. A., Farrell, J. M. & Stemwedel, M. E. (2000) Implicit motor sequence learning is represented in response locations. *Memory and Cognition* 28(3):366–75. [ASK]
Wills, S. J. & Mackintosh, N. J. (1999) Relational learning in pigeons? *Quarterly Journal of Experimental Psychology* 52B:31–52. [aPP]
Wolff, J. G. (1977) The discovery of segmentation in natural language. *British Journal of Psychology* 68:97–106. [MR]
Wolters, G. & Prinsen, A. (1997) Full versus divided attention and implicit memory performance. *Memory and Cognition* 25:764–71. [GW]
Wood, N. L., Stadler, M. A. & Cowan, N. (1997) Is there implicit memory without attention? A reexamination of task demands in Eich's (1984) procedure. *Memory and Cognition* 25:772–79. [aPP]
Woodworth, R. S. & Schlosberg, H. (1954) *Experimental psychology,* 3rd edition. Methuen. [DJM]
Wright, R. L. & Burton, A. M. (1995) Implicit learning of an invariant: Just say no. *Quarterly Journal of Experimental Psychology* 48A:783–96. [aPP]
Wulf, G. & Schmidt, R. A. (1997) Variability of practice and implicit motor learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 23:987–1006. [aPP]
Zajonc, R. B. (1980) Feeling and thinking: Preferences need no inferences. *American Psychologist* 35:151–75. [DH]
(2001) Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science* 10:224–28. [RFB]
Zbrodoff, N. J. (1999) Effects of counting in alphabet arithmetic: Opportunistic stopping and priming of intermediate steps. *Journal of Experimental Psychology: Learning, Memory and Cognition* 25:299–317. [aPP]
Zelazo, P. D. (1999) Language, levels of consciousness, and the development of intentional action. In: *Developing theories of intention: Social understanding and self-control,* ed. P. D. Zelazo, J. W. Astington & D. R. Olson. Erlbaum. [RAC]
Zimba, L. & Blake, R. (1983) Binocular rivalry and semantic processing: Out of sight, out of mind. *Journal of Experimental Psychology: Human Perception and Performance* 9:807–15. [aPP]


https://doi.org/10.1017/S0140525X02460062 Published online by Cambridge University Press