Implicit learning and consciousness:
A graded, dynamic perspective

Axel Cleeremans
Cognitive Science Research Unit
Université Libre de Bruxelles CP 122
Avenue F.-D. Roosevelt, 50
1050 Bruxelles — BELGIUM
Email: `axcleer@ulb.ac.be`

Luis Jiménez
Facultad de Psicología
Universidad de Santiago
15706 Santiago de Compostela — SPAIN
Email: `jimenez@usc.es`

May, 2001

1. Introduction

While the study of implicit learning is nothing new, the field as a whole has come to embody — over the last decade or so — ongoing questioning about three of the most fundamental debates in the cognitive sciences: The nature of consciousness, the nature of mental representation (in particular the difficult issue of abstraction), and the role of experience in shaping the cognitive system. Our main goal in this chapter is to offer a framework that attempts to integrate current thinking about these three issues in a way that specifically links consciousness with adaptation and learning. Our assumptions about this relationship are rooted in further assumptions about the nature of processing and of representation in cognitive systems. When considered together, we believe that these assumptions offer a new perspective on the relationships between conscious and unconscious processing and on the function of consciousness in cognitive systems.

To begin in a way that reflects the goals of this volume, we can ask the question: “What is implicit learning for?” In asking this question, one presupposes that implicit learning is a special process that can be distinguished from, say, explicit learning or, even more pointedly, from learning *tout court*. The most salient feature attributed to implicit learning is of course that it is *implicit*, by which most researchers in the area actually mean *unconscious*. Hence the question "What is implicit learning for?" is in fact a way of asking about the function of consciousness in learning that specifically assumes that conscious and unconscious learning have different functions. The central idea that we will develop in this chapter is that conscious and unconscious learning are actually two different expressions of a single set of constantly operating graded, dynamic processes of adaptation. While this position emphasizes that conscious and unconscious processing differ only in degree rather than in kind, it is nevertheless not incompatible with the notion that consciousness has specific functions in the cognitive economy.

Indeed, our main conclusion will be that the function of consciousness is to offer *flexible adaptive control* over behavior. By *adaptive* here, we do not mean simply the possibility for an agent to select one course of action among several possibilities. This, as dozens of computer programs routinely demonstrate, can be achieved without consciousness. Instead, we assume that genuine *flexibility* necessarily involves phenomenal consciousness (subjective experience), to the extent that successful adaptation in cognitive systems seems to make it mandatory that behavioral changes be based on the rewarding or punishing qualia they are associated with. There would be no point, for instance, in avoiding dangerous behavior were it not associated with *feelings* of danger. Learning is thus necessarily rooted, we believe, in the existence of at least some primitive ability for the cognitive agents to *experience* the consequences of their behavior and to recreate these experiences independently of action. These primitive experiences can then, through more elaborate learning and developmental processes, become integrated into increasingly complex structures that include representations of the *self*, that is, into a set of representations and processes that enable an agent to entertain a third-person perspective on itself, or, in other words, to look upon itself *as though* it were another agent. We surmise that any information-processing system that is sufficiently complex to make such processes possible should be characterized as conscious — albeit we may never find out unless this system exhibits the only sort of consciousness that we know of first-hand, that is, human consciousness. We will not discuss this important epistemic debate any further short of noting (1) that it actually is what the Turing Test is about (see French, 2000, for further discussion of the Turing Test), and (2) that it is perfectly possible to develop simulations of some behavior that successfully mimics adaptation without requiring qualia, but then, presumably, only at a level of description that would fail to pass more elaborate testing.

Our primary goal in this chapter will thus be to outline a novel framework with which to think about the relationships between learning and consciousness. In section 2, we propose to define learning as "a set of philogenetically advanced adaptation processes that critically depend on an evolved sensitivity to subjective experience so as to enable agents to afford flexible control over their actions in complex, unpredictable environments". We continue by discussing the implications of such a definition of learning on current debates about (1) the nature of phenomenal experience (section 3) and about (2) the functions of consciousness in cognitive systems (section 4). In section 5, we turn to an overview of our own proposal, and continue by briefly illustrating how our framework can be used to understand diverse phenomena in domains such as priming, implicit learning, automaticity and skill acquisition, or development (section 6). We conclude the chapter (section 7) by considering issues that the framework does not address. We should add that this chapter is by no means intended to offer a complete overview of all relevant phenomena and theories, but rather to convey the flavor of what we believe to be an alternative framework in which to consider some of the central issues in the domain of implicit learning.

2. Adaptation, adaptive changes, and learning

Mounting evidence suggests not only that the brain is far more plastic than previously thought, but also that the effects of learning can be tracked all the way down to the organization of local connectivity. To wit: Expert string players exhibit larger-than-normal areas of the somatosensory cortex dedicated to representing input from the fingering digits (Elbert et al., 1995). Likewise, not only is posterior hippocampus, — a region of the brain involved in episodic and spatial memory — enlarged in experienced taxi drivers compared to subjects who do not have extensive experience in memorizing complex maps, but the observed size differences further depend on the amount of driving experience (Maguire et al., 2000). There is also considerable evidence that the brain can recover in various flexible ways after trauma, and even suggestions that the very organization of the somatosensory cortex (the famous Penfied homonculus) depends on pre-natal sensory experience (Farah, 1998). More recently, suggestive evidence for neurogenesis was also found in humans (Eriksson et al., 1998) — a finding that overturned decades of unquestioned — but, as it turns out, erroneous — assumptions about the lack of regenerative cellular processes in the adult brain. These often spectacular findings all reassert that adaptation plays a fundamental role in cognition, and that its effects can be traced all the way down to the manner in which specific neural circuits are organized.

Given this plethora of new findings hinting that the brain constantly adapts to the environment that it is immersed in, what can we say about the relationships between learning and consciousness? Should we consider processes of adaptation in general to be distinct from processes of learning? Is it the case, as some authors contend (see Perruchet & Vinter, this volume; Shanks & St. John, 1994) that learning is always accompanied by conscious awareness? One can ask the question in another way: Why *should* behavior always be available to conscious control? It might seem particularly adaptive for complex organisms to be capable of behavior that does not require conscious control, for instance because behavior that does not require monitoring of any kind can be executed faster or more efficiently than behavior that does require such control. Reflexes such as withdrawing one's hand from a fire are good instances of behaviors that have presumably evolved to the point that they have been incorporated in the functional architecture of an organism's central nervous system and cannot be controlled any longer (or perhaps, only with extensive training on self-control techniques).

The relative accessibility of different actions to conscious awareness suggests that an important distinction between adaptation in general and learning is, precisely, the extent to which consciousness accompanies each. Learning, according to many standard definitions (e.g., Anderson & Memory, 1995; Klein, 1991; Tarpy, 1997), constitutes a subset of philogenetically advanced adaptation processes that are characteristic of so-called “cognitive systems”, and through which relatively permanent and generally adaptive changes in the behavior or dispositions of the organism arise as the result of their previous "experiences" with the environment in which they are immersed. From such a definition, it follows that the distinction between learning phenomena and the superordinate class of adaptation phenomena to which they belong depends on the “cognitive” status of the systems in which such learning occurs, and on the ability of these systems to enjoy a particular kind of sensitivity — "experience". Thus, however many reasons there might be to consider adaptation and learning as fundamentally rooted in the same mechanisms, we do not think that learning can simply be equated with adaptation. Adaptation, indeed, is a very broad concept. When taken to its limit, it might be used to refer to any dynamic relationship between an object and its environment through which (1) the object changes its states and dispositions (2) as a result of its prior sensitivity to the environment and (3) in a way that continuously modifies this sensitivity. It should be clear that by this definition, even inanimate objects such as rocks, thermostats or computer programs all exhibit patterns of adaptation. Indeed, erosion in rocks, the switch of a relay in a thermostat, or the occurrence of specific digital states in computers, can all be characterized as adaptive “responses” to changing environmental conditions, to the extent that they modify the systems’ future sensitivity to the reoccurrence of the same environmental conditions. In living systems, these processes of adaptation are further subject to continuous evolution on a species basis through the laws of natural selection.

Is it reasonable to consider such processes as processes of learning? Consider again standard definitions of learning. What, exactly, in these definitions, does “experience” refers to? Should our “experiences” as human beings be considered as similar to those of stones and amoebas? Certainly not. However, the literature about learning is in general conspicuously prone to conflate the term "experience" with any other kind of phenomenally neutral sensitivity that produces relatively permanent and adaptive changes in the responses of a system. For instance, even though neither machines nor neurovegetative systems are generally considered to be endowed with subjective experience, there is at least one journal that is entirely devoted to "Machine Learning". It is also relatively easy to find articles in psychological journals in which the changes produced in our neurovegetative systems in response to their environment are analyzed as examples of learning (e.g., Ader & Cohen, 1985).

While this conflation between “experience” and “mere sensitivity” has had the merit of emphasizing that there is a continuity between the processes of change that occur in different natural or artificial systems, it also blurs the distinction between learning and adaptation phenomena in general. In so doing, it has also further contributed to doing away with the distinction between cognitive and non-cognitive systems. Dennett (1996), in particular, has made this conflation completely explicit by assuming that the differences between cognitive and non-cognitive systems (e.g., between most animals and plants) might only be in the eye of the beholder. Indeed, according to Dennett, the main difference between animals and plants is that we tend to adopt an *intentional stance* when analyzing animals’ behavior, but do not do so when it comes to understand the dynamics of plant adaptation. As he boldly puts it, there is no reason to dispute the claim that plants should be considered as extremely slow animals whose "experiences" are overlooked because of our “temporal scale” chauvinism (Dennett, 1996), or that libraries should be taken as cognitive systems that use researchers as tools to reproduce themselves (Dennett, 1991).

While this conclusion strikes many of us as bluntly absurd, perhaps its absurdity should be taken as an indication that we need to revisit the notion of “experience” and, in so doing, attempt to carefully delineate what it entails. Indeed, if learning is a fundamental element of what it takes for a system to be “cognitive” (e.g., Dretske, 1988), it might also be the case that the nature of the phenomenal states upon which learning operates is essential to distinguish it from other processes of adaptation. This analysis thus forces us to look into the nature of phenomenal experience in some detail. That is what we attempt to do in the next section.

3. Consciousness

What is consciousness? While it would be foolish to even attempt to answer this question in this chapter, it might nevertheless be useful to offer guidelines about the sorts of explanations we are looking for, and about which of these are relevant to the study of implicit learning. In the following, we briefly discuss three aspects of consciousness that often tend to be overlooked in discussions of implicit learning: The fact that consciousness is not a unitary phenomenon, the fact that consciousness is graded, and the fact that consciousness is dynamic.

First, consciousness is not a unitary concept, but instead includes different dimensions. Block (1995), for instance, distinguishes between access consciousness, phenomenal consciousness, monitoring consciousness and self consciousness. Everybody agrees that the most problematic aspect of consciousness is phenomenal consciousness, or subjective experience, that is, the fact that information processing is accompanied by qualia — elements of conscious imagery, feelings or thoughts that together appear in our mind to form a coherent impression of the current state of affairs.

In the specific context of research about implicit learning, the central question is thus: Can changes in behavior occur without correlated changes in subjective experience, and are these changes best characterized as mere adaptation or as learning? This, at it turns out, is also one of the central questions in the ongoing "search for the neural correlates of consciousness" that has been the focus of so much recent empirical research about consciousness in the cognitive neurosciences. In an excellent overview, Frith, Perry and Lumer (1999) have suggested to organize paradigms through which to study the "neural correlates of consciousness" in nine groups resulting from crossing two dimensions: (1) three classes of psychological processes involving respectively knowledge of the past, present, and future — memory, perception, and action —, and (2) three types of cases where subjective experience is incongruent with the objective situation — cases where subjective experience fails to reflect changes in either (1) the stimulation or in (2) behavior, and (3) cases where subjective experience changes whereas stimulation and behavior remain constant.

The paradigmatic example of the third situation is binocular rivalry, in which an unchanging compound stimulus consisting of two elements each presented separately and simultaneously to each eye produces spontaneously alternating complete perceptions of each element. By asking participants to indicate which stimulus they perceive at any moment, one can then hope to establish which regions of the brain exhibits activity that correlates with subjective experience and which do not, in a situation where the actual stimulus remains unchanged. Frith et al. go on by delineating many other relevant empirical paradigms involving both normal subjects as well as patients suffering from a variety of neuropsychological syndromes. While reviewing these different paradigms in detail goes far beyond the scope of this chapter, it is interesting to note that implicit learning, in their analysis, constitutes one example of cases where subjective experience remains constant while behavior changes. The study of implicit learning is thus highly relevant to the study of consciousness in general.

In addition to the well-known difficult challenges involved in designing empirical paradigms suitable for the exploration of differences between conscious and unconscious processing (see Cleeremans, 1997 for an overview of these issues), the study of consciousness also notoriously involves a great deal of conceptual issues. In this respect, it is worth pointing out that current theories of consciousness indeed make sometimes very contrasted assumptions about its underlying mechanisms. For instance, Farah (1994) proposed to distinguish between three types of neuroscientific accounts of consciousness: "Privileged Role" accounts, "Integration" accounts, and "Quality of Representation" accounts. "Privileged Role" accounts take their roots in Descartes' thinking and assume that consciousness depends on the activity of specific brain systems whose function it is to produce subjective experience. "Integration" accounts, in contrast, assume that consciousness only depends on processes of integration through which the activity of different brain regions can be synchronized or made coherent so as to form the contents of subjective experience. Finally, "Quality of Representation" accounts assume that consciousness depends on particular properties of neural representations, such as their strength or stability in time.

In a recent overview article (Atkinson, Thomas, & Cleeremans, 2000, see also O'Brien & Opie, 1999), we proposed to organize computational theories of consciousness along two dimensions: (1) A process vs. representation dimension, which opposes models that characterize consciousness in terms of specific processes operating over mental representations, with models that characterize consciousness in terms of intrinsic properties of mental representations, and (2) A specialized vs. non-specialized dimension, which contrasts models that posit information-processing systems dedicated to consciousness with models for which consciousness can be associated with any information-processing system as long as this system has the relevant properties. Farah's three categories can be subsumed in this analysis in the following manner: "Privileged Role" models, which assume that some brain systems play a specific role in subtending consciousness, are specialized models that can be instantiated either through vehicle or through process principles. "Quality of Representation", models, on the other hand, are typical vehicle theories in that they emphasize that what makes some representations available to conscious experience are properties of those representations rather than their functional role. Finally, Farah's "Integration" models are examples of non-specialized theories, which can again be either instantiated in terms of the properties of the representations involved or in terms of the processes that engage these representations.

Atkinson et al.’s analysis thus offers four broad categories of computational accounts of consciousness:

(1) Specialized vehicle theories, which assume that consciousness depends on the properties of the representations that are located within a specialized system in the brain. An example of such accounts is Atkinson and Shiffrin’s (Atkinson & Shiffrin, 1971) model of short-term memory, which specifically assumes that representations contained in the short-term memory store (a specialized system) only become conscious if they are sufficiently strong (a property of representations).

(2) Specialized process theories, which assume that consciousness arises from specific computations that occur in a dedicated mechanism, as in Schacter’s CAS (Conscious Awareness System) model (Schacter, 1989). Shacter’s model indeed assumes that the CAS’s main function is to integrate inputs from various domain specific modules and to make this information available to executive systems. It is therefore as specialized model in that it assumes that there exist specific regions of the brain whose function it is to make its contents available to conscious awareness. It is a process model to the extent that any representation that enters the CAS will become available to conscious awareness in virtue of the processes that manipulate these representations, and not in virtue of properties of those representations themselves.

(3) Non-specialized vehicle theories include any model that posits that availability to consciousness only depends on properties of representations, regardless of where in the brain these representations exist. O’Brien & Opie’s “connectionist theory of phenomenal experience” (O'Brien & Opie, 1999) is the prototypical example of this category, to the extent that it specifically assumes that any stable neural representation will both be causally efficacious and form part of the contents of phenomenal experience.

(4) Non-specialized process theories, finally, are theories in which it is assumed that representations become conscious whenever they are engaged by certain specific processes, regardless of where these representations exist in the brain. Most recent proposals fall into this category. Examples include Tononi and Edelman’s “dynamic core” model (Tononi & Edelman, 1998); Crick and Koch’s idea that synchronous firing constitutes the primary mechanisms through which disparate representations become integrated as part of a unified conscious experience (Crick & Koch, 1995), or Grossberg’s characterization

> of consciousness as involving processes of “resonance” through which representations that simultaneously receive bottom-up and top-down activation become conscious because of their stability and strength (Grossberg, 1999).

While most recent neuro-computational models of consciousness fall into the last category, several proposals also tend to be somewhat more hybrid, instantiating features and ideas from several of the categories described by Atkinson et al. Baars' influential "Global Workspace" model (Baars, 1988), for instance, incorporates features from specialized process models as well as from non-specialized vehicles theories, to the extent that the model assumes that consciousness involves a specialized system (the global workspace), but also characterizes conscious states in terms of the properties associated with their representations (i.e. global influence and widespread availability) rather than in terms of the processes that operate on these representations. Likewise, Dehaene and Naccache’s recent “neural workspace” framework (Dehaene & Naccache, 2001) assumes that consciousness depends (1) on the existence of a distributed system of long-range connectivity that links many different specialized processing modules in the brain, and (2) on the simultaneous bottom-up and top-down activation of the representations contained in the linked modules. Thus, this model acknowledges both the existence of specific, dedicated mechanisms to support consciousness as well as the specific properties of representations (e.g., their strength or stability) brought about by specific processes (e.g., resonance).

These various tentative accounts of the neural or computational mechanisms of consciousness are highly relevant to the study of implicit learning because any theory of the mechanisms through which implicit learning occurs necessarily also has to make corresponding assumptions about the mechanisms of consciousness. As we shall see in section 4, however, most existing theories of implicit learning tend to be rather mute about their implications with respect to the study of consciousness. Indeed, most of the debate in the psychological literature about the relationships between conscious and unconscious processing has been dedicated to addressing methodological rather than conceptual issues. While these methodological debates are of central importance, we also believe that addressing the conceptual issues is essential.

A second central aspect of conscious experience — and one that is also particularly relevant for behavioral studies of implicit cognition, is that consciousness is not an all-or-none process or property, but that it affords many degrees and components. Conscious experience, however unified it appears to us, is not a single thing. Any theory of consciousness therefore has to answer questions about how the various elements of conscious experience are integrated with each other so as to form a unified whole, and about how to best think about the relative complexity of different sorts of conscious experiences. In other words: How does one go from the simple experiences that a snail might enjoy of its surroundings to the considerably more complex experience produced by your reading these words? How does one account for the differences between the sort of consciousness that infants undoubtedly possess to the sort of verbally rich consciousness that adults enjoy? Process and vehicle theories of consciousness make very different assumptions about these questions. For O'Brien & Opie (1999), for instance, the graded character of conscious experience is readily accommodated by vehicle theories, to the extent that properties of representations such as strength or stability in time can easily be mapped onto corresponding degrees or components of conscious awareness. This mapping is somewhat more delicate for what we have called process theories, even though at first sight they appear to offer an appealing set of conceptual principles with which to understand how conscious experience can increase in complexity through development or learning.

Dienes & Perner (1999) have recently pursued this goal in their theory of explicit and implicit knowledge, and "higher-order thought" (HOT) theories of consciousness in general can be described as relying on this principle (e.g., Rosenthal, 1986, 1997). However, what is harder to accept from such accounts of subjective experience is that its phenomenal character could be brought about in the first place from a series of computational processes performed on otherwise non-phenomenal representations. Indeed, and however much one might disagree with the specific way in which this thought experiment was framed, Searle's Chinese Room argument showed us twenty years ago that the phenomenal properties of experience just seem not to be the sort of stuff that one might expect to obtain by mere shuffling of formal representations or symbols, no matter how convoluted, recurrent, or complex the relation among these symbols may turn out to be (Searle, 1980; 1992; 1999). Neither semantics nor phenomenal experience can emerge out of syntax. Symbols need to be grounded. Hence, if this intuition is right, a pure process theory could never tell us the last word in accounting for the first principles of consciousness.

Vehicle theories, it therefore appears, appear to be the best candidates to account for the emergence of the first elements of subjective experience which, through processes of learning, development and socialization, subsequently provide the appropriate foundations for the emergence of more elaborated forms of consciousness. It must be made clear at this point that by "vehicle theories" we refer to any theory that assumes that experience *is not* merely a relational or syntactic property that could be realized through any representational vehicle, but that claim instead that experience arises in a specific medium (e.g., neural) and as a result of processes that are proper to this medium[1].

For the sake of discussion, let us simply accept that phenomenal experience arises as the result of some neural processes. What, then, might be the functions fulfilled by phenomenal experience? What is it about experience that makes it play a special role in distinguishing between learning and mere adaptation? These questions are in fact questions about a third aspect of consciousness, that is, its dynamical character. Most discussions of consciousness tend to analyze it as a static property of some processes or representational states. However, it is obvious that consciousness is a phenomenon that is highly dynamical: What I am aware now I might be unaware of at the next moment. Likewise, what I am aware of at some point in time when learning a new skill is not identical with what I will be aware of after I have mastered the skill. Thus, we therefore believe that processes of change are central to our understanding of consciousness, and that an analysis of its possible functions should therefore be rooted in an analysis of the role that learning and adaptation play in shaping action.

4. The function of consciousness: Commander Data meets the Zombies

The findings briefly overviewed at the beginning of section 2 raise the question of what the role of consciousness might be in adaptation and learning. We concluded that a significant difference between adaptation and learning is whether or not consciousness is involved. In this section, we attempt to reflect upon the function that consciousness might have in information processing. In so doing, we suggest that most existing theories of the relationships between conscious and unconscious processing have simply failed to give consciousness a clear functional role.

In a recent overview article, Dehaene and Naccache (2001) conclude that “The present view associates consciousness with a unified neural workspace through which many processes can communicate. The evolutionary advantages that this system confers to the organism may be related to the increased independence that it affords.” (p. 31). Dehaene and Naccache thus suggest that consciousness allows organisms to free themselves from acting out their intentions in the real world, relying instead on less hazardous simulation made possible by the neural workspace. While we certainly agree with this conclusion, it begs the question of how consciousness came to play these functions in the first place. Are there any adaptive or evolutionary causes that would favor the emergence of unifying control systems characterized by conscious states, and that could go beyond what local adaptive processes can do by forcing large parts of the nervous system to work together in a coherent direction for some fractions of seconds? How can these coherent, resonant, synchronous, reverberant, or otherwise conscious states of the system come to reflect the most adaptive representation of the current situation, given that "what is most adaptive" continuously changes?

As discussed by Perruchet and Vinter (this volume), the answers to these questions are intimately related to the dynamics between learning and consciousness: On the one hand, phenomenal consciousness provides the cognitive system with an adapted, global representation of the current situation so that learning mechanisms operate on the best possible representations. On the other hand, learning changes these representation in increasingly adaptive ways. From this perspective then, the central function of consciousness is to offer flexible, adaptive control over behavior.

This complex, dynamical relationship between consciousness and learning has, however, often tended to be overlooked in classical models of cognition. As argued in Cleeremans (1997) and also in Jiménez and Cleeremans (1999), this is most likely due to the fact that classical models of cognition (the "Computational Theory of Mind", see Fodor, 1975) take it as a starting point that cognition is symbol manipulation. As we will try to highlight in the next few paragraphs, we surmise that one takes cognition to be exclusively and exhaustively about symbol manipulation, then there are but a few conceptual possibilities with which to think about differences between conscious and unconscious states.

Cognitive scientists concerned with the relation between consciousness and cognition generally tend to oscillate between two extreme (and admittedly caricatural) positions, which we have dubbed "Commander Data" and "Zombie" theories of cognition. Star Trek's character Data is an android whose bodily and cognitive innards are fully transparent to himself. Except in rare circumstances (which systematically tend to be described as the result of some sort of dysfunction), Data is thus capable of describing in uncanny detail each and every aspect of its internal states: How much force he is applying when attempting to pry open a steel door, how many circuits are currently active in his positronic brain, or the number of times over the last ten years he smelled a particular scent, and in which circumstances he did so, etc. Commander Data theorists likewise assume that cognition is fully transparent, that is, (1) that whatever knowledge is expressed through behavior is also transparently available to introspection, and (2) that consciousness reigns supreme and allows access, with sufficient effort or attention, to all aspects of our inner lives. This perspective is what Broadbent described as the “common sense” view of cognition, according to which “people act by consulting an internal model of the world, a database of knowledge common to all output processes, and manipulating it to decide on the best action” (Broadbent, Fitzgerald, & Broadbent, 1986, p. 77).

In contrast, the famed philosophical zombies (Chalmers, 1996) are perfectly opaque, and in this sense instantiate absolutely implicit beings: Whatever internal knowledge currently influences their behavior can neither be explicit nor conscious because, by definition, they lack conscious experience. Zombie theorists thus take it as a starting point that consciousness has an epiphenomenal character: There is a zombie within you and, while you may not be aware of its existence, it could in fact be responsible for most of your actions. It is capable of processing all the information you can process in the same way that you do, with one crucial difference: “All is dark inside” (Chalmers, 1996, p. 96); your zombie is unconscious. From this perspective then, cognition is inherently opaque, and consciousness, when present, offers but a very incomplete and imperfect perspective on internal states of affairs.

Needless to say, both of these perspectives are profoundly unsatisfactory. On the one hand, Zombie perspectives (ZP) ascribe no role whatsoever to consciousness in information processing, threaten to rob us of free will, and — because it is absurd to deny consciousness altogether — are ultimately forced to assume the existence of equally powerful conscious and unconscious systems. On the other hand, Commander Data perspectives (CDP), by assuming that all of cognition is conscious, paradoxically likewise depict consciousness as epiphenomenal. Crucially, both perspectives assume that consciousness does not change cognition in any principled way, and hence that consciousness plays no functional role beyond that of a epiphenomenon that accompanies either a functionally redundant subset of (ZP) or all (CDP) cognitive events.

On the face of the deeply counterintuitive flavor of both perspectives, it seems surprising to see that the past few years have witnessed the appearance of several broad theoretical proposals that intentionally or inadvertently endorse either of these perspectives. Some of these proposals are based on empirical evidence, and argue that there is in fact no evidence for unconscious influences in cognition. Thus for instance, Holender (1986), based on an extensive review of the subliminal perception literature, found no evidence for the existence of unconscious priming. Holender (1992) further proposed that many congruency effects observed in priming experiments can be accounted by conflicts between conscious contents, that is, without appealing to the effects of unconscious influences. Likewise, Shanks and StJohn (1994), expanding on the perspective offered by Brewer (1976), concluded their target article dedicated to implicit learning by the statement that “Human learning is almost invariably accompanied by conscious awareness” (p. 394).

Other proposals are more conceptual in nature. For instance, O’Brien and Opie (O'Brien & Opie, 1999) propose that the contents of phenomenal consciousness include all stable neural states, and that it is only those stable states that are “causally efficacious”, that is, susceptible to influence further processing and, ultimately, behavior. Perruchet and Vinter (1998, this volume), consider that unconscious influences on behavior should be ascribed exclusively to noncognitive, neural processes and state that “Mental life […] is co-extensive with consciousness” (Perruchet, personal communication, see also Dulany, 1997). Finally, Dienes and Perner’s (1999) recent “theory of implicit and explicit knowledge”, while carefully delineating the various ways in which knowledge can be cast as implicit or explicit, also seems to take it as a starting point that causally efficacious knowledge is always explicit in some sense, that is, at least at the specific level that is needed to account for the observed behavioral effects, and hence ends up, we believe, inadvertently painting a picture of cognition in which the implicit again plays no functional role in cognition.

It should be pointed out that if the emphasis of these theories on the "transparent" character of cognition can be seen as a normal swing of the conceptual pendulum, there is nevertheless something paradoxical about the emergence of such proposals at a time when the importance of unconscious processing in cognition finally appears to have gained some form of recognition in dozens of articles, books and conferences.

The debate, we believe, is not so much rooted in equivocal empirical findings, but rather in the deep conceptual problems associated with the notion of unconscious representation. Hence, defenders of the claim that cognition can be unconscious often succumb to some version of the ZP, while defenders of the opposite view can often be taken to endorse some variant of the CDP. Crucially, we believe that both these general frameworks are in fact based on the classical assumption that cognition involves symbol manipulation, and hence that their only way to separate conscious from unconscious cognition is to assume that unconscious cognition is just like conscious cognition, but only minus consciousness (Searle, 1992).

In the next section, we would like to sketch out an alternative, subsymbolic, framework through which to think about the relationship between learning and consciousness — one that we believe offers a clear function to consciousness by linking it with adaptability in cognitive systems, while at the same time leaving open the possibility for adaptive changes to occur without consciousness.

5. The framework

If our central assumption that the function of consciousness is to offer adaptive control over behavior is correct, then consciousness is necessarily closely related to processes of learning, because one of the central consequences of successful adaptation is that conscious control is no longer required over the corresponding behavior. We therefore believe that it makes sense to root accounts of consciousness in accounts of how change occurs in cognitive systems.

Like Perruchet & Vinter (this volume), we assume that there is a dynamic relationship between consciousness and learning such that (1) awareness of a particular state of affairs triggers learning and (2) that this learning in turn changes the contents of subjective experience so as to make these contents more adapted. However, and this is an important departure from Perruchet & Vinter's framework, we also assume that learning has additional obligatory indirect effects that can fail to enter awareness. In other words, learning is not *just* about modifying conscious experience, as Perruchet & Vinter seem to assume. Thus, when I learn about cats, I also indirectly learn about dogs and other animals, because the corresponding representations are all linked together by virtue of being embedded in distributed representational systems. These indirect effects of conscious learning need not themselves be conscious, particularly if they are weak.

We will return to these issues in the general discussion. At this point, we would like to introduce the set of assumptions that together form our framework. In the following, we present these assumptions in four groups: Assumptions about information processing (P1-4), about representation (R1-3), about learning (L1-3) and, finally, about consciousness (C1-5).

5.1 Assumptions about information processing

Consistently with well-known ideas in the connectionist literature (e.g., Rumelhart & McClelland, 1986), we will assume the following without further discussion:

P1. *The cognitive system is best viewed as involving a large set of interconnected processing modules organized in a loose hierarchy. Each module in turn consists of a large number of simple processing units connected together.*

P2. *Long-term knowledge in such systems is embodied in the pattern of connectivity between the processing units of each module and between the modules themselves.*

P3. *Dynamic, transient patterns of activation over the units of each module capture the results of information processing conducted so far.*

P4. *Processing is graded and continuous: Connected modules continuously influence each other's processing in a graded manner that depends on the strength of the connection between them and on the strength of the activation patterns that they contain.*

5.2. Assumptions about representation

Representation is one of the most difficult issues to think about in the cognitive sciences because it is often delicate to delineate exactly which states should be properly taken to be representational (see Dienes & Perner, this volume, for a detailed discussion of representation). In the following, and in contrast to purely dynamical approaches, we take the perspective that representations are necessary as mediating states through which the intermediate results of processing can be captured, thereby making it possible for complex tasks to be decomposed into modular components.

R1. *Representations consist exclusively of the transient patterns of activation that occur in distributed memory systems*

This assumption is a central one in our framework because it contrasts with other recent proposals (e.g., Dienes & Perner, this volume). In particular, we do not think that the knowledge that is embedded in the pattern of connectivity between units of a module or between modules themselves is representational in the same manner that patterns of activation are. Indeed, while such knowledge can be analyzed as representational from a third-person perspective (because the connection between two units, for instance, can be described as representing the fact that the units' activity are correlated), it is never directly available to the system itself. In other words, such knowledge is knowledge "in the system" rather "for the system" (see Clark & Karmiloff-Smith, 1993). Knowledge embedded in connections weights can thus only be expressed dynamically, over the course of some processing, when the corresponding representations form over a given set of processing units. These representations can then in turn influence further processing in other modules. Importantly, and in contrast to thoroughly classical approaches in cognitive science, the extent to which representations can influence processing in such systems never depends on representations being interpreted by a “processor”.

R2. *Representations are graded: They vary on several dimensions that include strength, stability in time, and distinctiveness*

Patterns of activation in neural networks and in the brain are typically distributed and can therefore vary on a number of dimensions, such as their stability in time, their strength. or their distinctiveness. Stability in time refers to how long a representation can be maintained active during processing. There are many indications that different neural systems involve representations that differ along this dimension. For instance, the prefrontral cortex, which plays a central role in working memory, is widely assumed to involve circuits specialized in the formation of the enduring representations needed for the active maintenance of task-relevant information. Strength of representation simply refers to how many processing units are involved in the representation, and to how strongly activated these units are. As a rule, strong activation patterns will exert more influence on ongoing processing than weak patterns. Finally, distinctiveness of representation refers to the extent of overlap that exists between representations of similar instances. Distinctiveness has been hypothesized as the main dimension through which cortical and hippocampal representations differ (McClelland, McNaughton, & O'Reilly, 1995; O'Reilly & Munakata, 2000), with the latter becoming active only when the specific conjunctions of features that they code for are active themselves.

In the following, we will collectively refer to these different dimensions as “quality of representation” (see also Farah, 1994) For our purposes, the most important notion that underpins these different dimensions is that representations, in contrast to the all-or-none prepositional representations typically used in classical theories, instead have a graded character which enables any particular representation to convey in a natural manner the extent to which what it refers to is indeed present. A second important aspect of this characterization of representational systems in the brain is that representations are complex, distributed objects that systematically tend to involve many processing units.

R3. *Representations are dynamic, active, and constantly causally efficacious.*

This assumption simply states that memory traces, far from being static propositions waiting to be accessed by some process, instead continuously influence processing regardless of their strength, stability, or distinctiveness. This assumption is again central in any connectionist account of cognition. Indeed, it takes its roots in McClelland's analysis of cascaded processing (McClelland, 1979), which, by showing how modules interacting with each other need not "wait" for other modules to have completed their processing before starting their own, demonstrated how stage-like performance could emerge out of such continuous, non-linear systems. Thus, even weak, poor-quality traces, in our framework are capable of influencing processing, for instance through associative priming mechanisms, that is, in conjunction with other sources of stimulation. Strong, high-quality traces, in contrast have generative capacity, in the sense that they can influence performance (i.e., determine responses) independently of the influence of other constraints, that is, whenever their preferred stimulus is present.

5.3. Assumptions about learning

Having put in place our assumptions about processing and representation, we now focus on learning mechanisms. We assume the following:

L1. *Adaptation is a mandatory consequence of information processing*

Every form of neural information processing produces adaptive changes in the connectivity of the system, through mechanisms such as Long Term Potentiation (LTP) or Long Term Depression (LTD) in neural systems, or hebbian learning in connectionist systems. An important aspect of these mechanisms is that they are mandatory in the sense that they take place whenever the sending and receiving units or processing modules are co-active. O'Reilly and Munakata (2000) have described hebbian learning as instantiating what they call model learning. The fundamental computational objective of such unsupervised learning mechanisms is to enable the cognitive system to develop useful, informative models of the world by capturing its correlational structure. As such, they stand in contrast with task learning mechanisms, which instantiate the different computational objective of mastering specific input-output mappings (i.e., achieving specific goals) in the context of specific tasks through error-correcting learning procedures. Regardless of how these two classes of learning mechanisms can be combined, the important point to remember in the context of this framework is that model learning operates whenever information processing takes place, whereas task learning only operates in specific contexts defined by particular goals.

L2. *Learning is adaptation that specifically involves high-quality representations.*

We assume that learning consists specifically of those adaptation processes that involve high-quality, strong, stable representations. One way to characterize this notion is to appeal to another distinction offered by O’Reilly & Munakata (2000) — that between weight-based and activation-based processing. According to O’Reilly & Munakata, “Activation-based processing is based on the activation, maintenance, and updating of active representations to influence processing, whereas weight-based processing is based on the adaptation of weight values to alter input/output mappings” (p. 380). The main advantage of activation-based processing is that it is faster and more flexible than weight-based processing. Speed and flexibility are both salient characteristics of high-level cognition. O’Reilly & Munakata further speculate that activation-based processing is one of the central characteristics of the frontal cortex, and suggest that this region of the brain has evolved specifically to serve a number of important functions related to controlled processing, such as working memory, inhibition, executive control, and monitoring or evaluation of ongoing behavior. To serve these functions, processing in the frontal cortex is characterized by mechanisms of active maintenance through which representations can remain strongly activated for long periods of time so as it make it possible for these representations to bias processing elsewhere in the brain.

O’Reilly and Munakata point out that a major puzzle is to understand how the frontal cortex comes to develop what they call a “rich vocabulary of frontal activation-based processing representations with appropriate associations to corresponding posterior-cortical representations” (p. 382). Our framework does not solve this difficult chicken-and-egg problem, but simply suggests that early learning or development, which involve mostly weight-based processing, progressively results in the emergence of the strong, high-quality representations that allow activation-based processing and the ensuing flexibility to take place. Language and linguistic representations in general undoubtedly play a major role in making activation-based processing possible.

L3. *Learning has both direct and indirect effects.*

Learning not only has direct effects, (i.e., changing the subjective experience that corresponds to the processing of a particular event and modifying the system’s response to that event), but it also has indirect effects on how (functionally or physically) similar events are processed. This is again a natural consequence of the assumption that memory systems in general involve distributed, superpositional representations, such that all representations share many processing units, and such that all processing units are involved in many representations. In such representational systems, changes to any particular representation that might arise from learning will necessarily have indirect effects on related representations. Importantly, these indirect effects are mediated by changes in the connection weights shared by the different representations in a given module; in other words, they do not involve direct, simultaneous modification of the corresponding representations. These indirect effects are thus, in our framework, not necessarily accompanied by awareness, because to be accompanied by awareness, their origin and magnitude would have to be identifiable by the agent.

5.4 Assumptions about consciousness

So far, we have spelled out a number of assumptions about information processing, representation, and learning. We are now ready to introduce our assumptions about consciousness and its relationship to adaptation and learning processes. The central ideas that we would like to explore are (1) that the extent to which a particular representation is available to consciousness depends on its quality, (2) that learning produces, over time, higher-quality, adapted representations, and (3) that the function of consciousness is to offer necessary control over those representations that are strong enough to influence behavior, yet not sufficiently adapted that their influence does not require control anymore.

Insert Figure 1 about here

Figure 1 aims to capture these ideas by representing the relationships between quality of representation (X-axis) on the one hand and (1) potency, (2) availability to control, (3) availability to subjective experience. We discuss the figure at length in the following section. Let us simply note here that the X-axis represents a continuum between weak, poor-quality representations on the left and very strong, high-quality representations on the right., and that principle R3 (“Representations are constantly causally efficacious”) is captured by the curve labeled “potency”, which assumes that all representations, even weak ones, can influence behavior to some extent. The general form of the relationship between quality of representation and potency is assumed to be non-linear.

Two further points are important to keep in mind with respect to Figure 1. First, the relationships depicted in the Figure are intended to represent availability to some dimension of behavior or consciousness independently of other considerations. Many potentially important modulatory influences on the state of any particular module are thus simply not meant to be captured neither by Figure 1, nor by our framework as we present it here. Second, the figure is intended to represent what happens in each of many processing modules involved in any particular cognitive task. Thus, as hinted by assumptions P1-P4, at any point in time, there will be many such modules active, each contributing to some extent to behavior and to conscious experience; each modulating the activity of other modules. With these caveats in mind, let us now turn to our five assumptions about consciousness and its relationship with learning:

C1. *Consciousness involves two dimensions: Subjective experience and control*

As argued by many, and most cogently by Ned Block, consciousness involves at least two separable aspects, namely access consciousness (A-consciousness) and phenomenal consciousness (P-consciousness). For Posner and Rothbart (Posner & Rothbart, 1998), awareness of the sensory world and voluntary control are the two most important aspects of consciousness. According to Block (1995), “A perceptual state is access-conscious roughly speaking if its content — what is represented by the perceptual state — is processed via that information processing function, that is, if its content gets to the Executive system, whereby it can be used to control reasoning and behavior.” (p. 234). In other words, whether a state is A-conscious is defined essentially by the causal efficacy of that state; the extent to which it is available for global control of action. Control refers to the ability of an agent to control, to modulate, and to inhibit the influence of particular representations on processing. In our framework, control is simply a function of potency, as described in assumption C3. In contrast, P-consciousness refers to the phenomenal aspects of subjective experience: A state is P-conscious to the extent that there is something it is like to be in that state. While the extent to which potency (i.e., availability to access consciousness), control, and subjective experience (i.e., availability to phenomenal consciousness) are dissociable is debatable, our framework suggests that these three aspects of consciousness are closely related to each other.

C2. *Availability to consciousness correlates with quality of representation*

This assumption is also a central one in our framework. It states that explicit, conscious knowledge involves higher quality memory traces than implicit knowledge. "Quality of representation", as discussed above (assumption R3), designates several properties of memory traces, such as their relative strength in the relevant information-processing pathways, their distinctiveness, or their stability in time. Our assumption is consistent with the theoretical positions expressed by several different authors over the last few years. O’Brien & Opie

(1999) have perhaps been the most direct in endorsing a characterization of phenomenal consciousness in terms of the properties of mental representations in defending the idea that “consciousness equals stability of representation”, that is, that the particular mental contents that one is aware of at some point in time correspond to those representations that are sufficiently stable in time. Mathis & Mozer (1996) have also suggested that consciousness involves stable representations, but have defined stability more technically than O’Brien & Opie have, specifically by offering a computational model of priming phenomena in which stability literally corresponds to the state that a so-called dynamic “attractor” network reaches when the activations of a subset of its units stops changing and settle into a stable, unchanging state.

A slightly different perspective on the notion of “quality of representation” is offered by authors who emphasize not stability, but strength of representation as the important feature through which to characterize availability to consciousness. One finds echoes of this position in the writings of Kinsbourne (1997), for whom availability to consciousness depends on properties of representations such as duration, activation, or congruence. Importantly, for both O’Brien & Opie and for Kinsbourne, the contents of subjective experience never depend on representations entering a particular system in the brain — that is, consciousness is conceived as essentially decentralized: Any region of the brain can contribute to the contents of subjective experience so long as its representational vehicles have the appropriate properties.

In Figure 1, we have represented the extent to which a given representation is available to the different components of consciousness (phenomenal consciousness, access-consciousness/potency, and control) as functions of a single underlying dimension expressed in terms of the quality of this representation. Availability to access-consciousness is represented by the curve labeled “potency”, which expresses the extent to which representations can influence behavior as a function of their quality. We simply assume that high-quality, strong, distinctive representations are more potent than weaker representations. “Availability to control processes” is represented by a second curve, so labeled. We simply assume that both weak and very strong representations are difficult to control, and that maximal control can be achieved on representations that are strong enough that they can begin to influence behavior in significant ways, yet not so strong that have become utterly dominant in processing. Finally, availability to phenomenal experience is represented by the third curve, obtained by convolving the other two. The underlying intuition, discussed in the context of assumption C4, is that which contents enter subjective experience is a function of both availability to control and of potency.

C3. *Developing high-quality representations takes time*

This assumption states that the emergence of high quality representations (see assumption C2) in a given processing module takes time, both over training or development, as well as during processing of a single event. Figure 1 can thus be viewed as representing not only the relationships between quality of representation and their availability to the different components of consciousness, but also as a depiction of the dynamics of how a particular representation will change over the different time scales corresponding to development, learning, or within-trial processing.

Both skill acquisition and development, for instance, involve the long-term progressive emergence of high-quality, strong memory traces based on early availability of weaker traces. Likewise, the extent to which memory traces can influence performance at any moment (e.g., during a single trial) depends both on available processing time, as well as on overall trace strength. We envision these processes of change as operating on the connection weights between units in a connectionist network. They can involve either task-dependent, error-correcting procedures, or unsupervised procedures such as hebbian learning. In either case, continued exposure to exemplars of the domain will result in the development of increasingly congruent and strong internal representations that capture more and more of the relevant variance. Although we think of this process as essentially continuous, we distinguish three stages in the formation of such internal representations (each depicted as separate regions in Figure 1): Implicit representations, explicit representations, and automatic representations.

The first region, labeled “implicit cognition” in Figure 1, is meant to correspond to the point at which processing starts in the context of a single trial, or to some early stage of development or skill acquisition. In either case, this stage is characterized by weak, poor-quality representations. A first important point, embodied in assumption R3, is that representations at this stage are already capable of influencing performance, as long as they can be brought to bear on processing together with other sources of constraints, that is, essentially through mechanisms of associative priming and constraint satisfaction. A second important point is that this influence is best described as "implicit", because the relevant representations are too weak (i.e., not distinctive enough) for the system as a whole to be capable of exerting control over them: You cannot control what you cannot identify as distinct from something else. One might even speculate that what enables you to take control of an internal state is precisely the fact that it is capable of triggering responses in and of itself — a speculation that links control with action in a very direct way.

The second region of Figure 1 corresponds to the emergence of explicit representations, defined as representations over which one can exert control. In the terminology of attractor networks, this would correspond to a stage during learning at which attractors become better defined — deeper, wider, and more distinctive. It is also at this point that the relevant representations acquire generative capacity, in the sense that they now have accrued sufficient strength to have the potential to determine appropriate responses when their preferred stimulus is presented to the system alone. Because awareness is partially tied to control in our framework, one would thus also be aware both of these internal representations and of their influence on our behavior. Because one is aware of these representations, one can then also possess metaknowledge about them, and recode them in various different ways, for instance, as linguistic propositions.

The third region involves what we call automatic representations, that is, representations that have become so strong that their influence on behavior can not longer be controlled (i.e., inhibited). Such representations exert a mandatory influence on processing. Importantly, however, one is aware both of possessing them (that is, one has relevant metaknowledge) and of their influence on processing (see also Tzelgov, 1997), because availability to conscious awareness depends on the quality of internal representations, and that strong representations are of high quality. In this perspective then, one can always be conscious of automatic behavior, but not necessarily with the possibility of control over these behaviors.

In our framework, skill acquisition, and development therefore involve a continuum at both ends of which control over representations is impossible or difficult, but for very different reasons: Implicit representations influence performance but cannot be controlled because they are not yet sufficiently distinctive and strong for the system to even know it possesses them. This might in turn be related to the fact that, precisely because of their weakness, implicit representations cannot influence behavior on their own, but only in conjunction with other sources of constraints. Automatic representations, on the other hand, cannot be controlled because they are too strong, but the system is aware both of their presence and of their influence on performance.

C4. *The function of consciousness is to offer flexible, adaptive control over behavior*

Our framework gives consciousness a central place in information processing, in the sense that its function is to enable flexible control over behavior. Crucially, however, consciousness is not necessary for information processing, or for adaptation in general, thus giving a place for implicit learning in cognition. We believe this perspective to be congruent with theories of adaptation and optimality in general.

Indeed, another way to think about the role of learning in consciousness is to ask: "When does one need control over behavior?". Control is perhaps not necessary for implicit representations, for their influence on behavior is necessarily weak (in virtue of the fact that precisely because they are weak, such representations are unlikely to be detrimental to the organism even if they are not particularly well-adapted). Likewise, control is not necessary for automatic representations, because presumably, those representations that have become automatic after extensive training should be adapted (optimal) as long as the processes of learning that have produced them can themselves be assumed to be adaptive. Automatic behavior is thus necessarily optimal behavior, except, precisely, in cases such as addiction or in laboratory situations where the automatic response is manipulated to be non-optimal, such as in the Stroop situation. Referring again to Figure 1, this analysis thus suggests that the representations that require control are the explicit representations that correspond to the central region of Figure 1: Representations that are strong enough that they have the potential to influence behavior in and of themselves (and hence that one should really care about, in contrast to implicit representations), but not sufficiently strong that they can be assumed to be already adapted, as is the case for automatic representations. It is for those representations that control is needed, and, for this reason, it is of these representations that one is most aware of.

Likewise, this analysis also predicts that the dominant contents of subjective experience at any point in time consists precisely of those representations that are strong enough that they can influence behavior yet weak enough that they still require control. Figure 1 reflects these ideas by suggesting that the contents of phenomenal experience depend both on the potency of currently active representations as well as on their availability to control. Since availability to control is inversely related to potency for representations associated with automatic behavior, this indeed predicts weaker availability to phenomenal experience of “very strong” representations as compared to “merely strong” representations. Such "automatic representations" therefore form what Mangan (1993) has called the "fringe of consciousness". In other words, such representations can become conscious if appropriate attention is directed towards their contents — as in cases where normally automatic behavior (such as walking) suddendly becomes conscious because the normal unfolding of the behavior has been interrupted (e.g., because I've stumbled upon something) — but they are not normally part of the central focus of awareness nor do they require conscious control.

While the dominant contents of subjective experience can thus be viewed as reflecting the activity of the topmost module in the constantly evolving loose hierarchy of processing modules involved in any particular aspect of information processing, it is also important to note that we assume, in contrast to the position expressed by Perruchet & Vinter, that complex representations depend on the continued activation of their more elementary components. In other words, while learning certainly results in the elaboration of progressively more complex representations, it neither prevents their components from contributing to subjective experience nor does it eliminate their influence on ongoing processing. This therefore opens the door for the continued — but attenuated, indirect — expression of the representations associated with these lower-level modules.

C5. *Learning shapes conscious experience*

This assumption, which we adapt from Perruchet & Vinter (this volume) is a corollary of assumption C4: If the function of consciousness is to offer flexible, adaptive control over behavior, then its contents — the way it reflects the world — should necessarily be shaped by learning so that, at any moment, these contents tend to reflect precisely those aspects of the situation that most require control. This assumption allows us to relate two central aspects of consciousness that have often been considered as independent: subjective experience and control of action —or phenomenal vs. access consciousness.

5.5. Ways for knowledge to be implicit

In our framework, we emphasized quality of representation as a central dimension through which to account for which representations are likely to enter conscious awareness. It should be clear, however, that we take quality of representation as a necessary, but not sufficient condition, for conscious awareness. In particular, our framework remains mute with respect to the fate of the high-quality, strong representations that characterize explicit, conscious cognition, short of claiming that it is these representations that are most likely to form the contents of subjective experience. Whether these representations actually enter conscious experience is yet a different story — one in which processes of attention and processes of integration undoubtedly play a central role. In this respect, our framework is not inconsistent with recent proposals that emphasize the importance of such processes in the formation of subjective experience. One such recent proposal has been put forward by Dehaene and Naccache (2001). These authors, based on Baars' notion of "global workspace" (Baars, 1988), propose that conscious awareness depends on the extent to which the contents of the many domain-specific unconscious processing modules that make up our brain can be made accessible globally through specific, dedicated, long-distance neural pathways that interconnect the modules and specific regions of the brain (i.e., essentially prefrontal cortex, anterior cingulate and other regions connected to both). Availability to the global workspace thus depends on both bottom-up (i.e., input strength) and top-down (i.e. attention) factors. When these two conditions exist, the contents of those modules that connect to the neural workspace would then enter in the stable, resonant, or synchronous states that are assumed to correlate with conscious awareness.

Kanwisher (2001) also discusses the conditions under which particular representations will enter conscious awareness, and notes that activation strength alone, while perhaps necessarily, is certainly not sufficient. Like Dehaene and Naccache, Kanwisher suggests that awareness also depends on "informational access", that is, on the fact that other parts of the brain/mind have access to the relevant representations. Kanwisher also suggests the accessibility can change over time as a result of practice — a point that we very much agree with —, and that an important further factor in determining availability to consciousness is what she calls the "type/token" distinction, that is, the fact that awareness of some perceptual attribute not only requires a strong corresponding representation, but also "individuation of that perceptual information as a distinct event" (p. 107). In other words, the relevant representation has to be accompanied by relevant metaknowledge — a point discussed in detail by Dienes and Perner (1999).

Our own framework leaves open four distinct possibilities for knowledge to be implicit.

First, we assume that the knowledge that is embedded in the connection weights within and between processing modules can never be directly available to conscious awareness and control. This is simply a consequence of the fact that we assume that consciousness necessarily involves representations (patterns of activation over processing units). Because weight-based knowledge is not representational in this specific sense, it follows that it can never directly contribute to the contents of conscious experience. This knowledge will, however, shape the representations that depend on it, and its effects will therefore detectable — but only indirectly, and only to the extent that these effects are sufficiently marked in the corresponding representations.

Second, we assume that to enter conscious awareness, a representation needs to be a sufficiently high-quality in terms of strength, stability in time, or distinctiveness. Weak representations are therefore poor candidates to enter conscious awareness. This, however, as we repeatedly emphasized, does not necessarily imply that they remain causally inert, for they can influence further processing in other modules, even if only weakly so. Note that this aspect of our framework differs both from the assumptions put forward by O'Brien and Opie (1999) and from those embodied in Perruchet and collaborators (Perruchet, Vinter, & Gallego, 1997; Perruchet & Vinter, 1998; this volume).

Third, a representation can be strong enough to enter conscious awareness, but failed to be recognized as relevant to the particular situation that is currently unfolding. This case corresponds almost exactly with Kanwisher's "type/token" distinction, and also with aspects of Dienes & Perner's analysis of the differences between implicit and explicit knowledge. Conscious contents, indeed, have to be linked together in a coherent manner before they can be made available globally for conscious report and for the control of action. One should therefore be very careful in distinguish between cases involving awareness of the intention of initiating some behavior, awareness of the fact that the behavior is taking place, awareness of the causes of the behavior, and awareness of the effects of the behavior. There are thus many opportunities for a particular conscious content to remain, in a way, implicit, not because its representational vehicle does not have the appropriate properties, but because it fails to be integrated with other conscious contents. Dienes & Perner (this volume) offer an insightful analysis of the different ways in which what we call high-quality representations can remain implicit.

Finally, a representation can be so strong that its influence can be no longer be controlled. In theses cases, it is debatable whether the knowledge should be taken as genuinely unconscious, because they certainly can become fully conscious as long as appropriate attention is directed to them, but the point is that such very strong representations can trigger and support behavior without conscious intention and without the need for conscious monitoring of the unfolding behavior.

Another way to think about these different ways for knowledge to be implicit is to consider the various mechanisms of change suggested by O'Reilly & Munakata. Recall that these authors distinguish between weight-based processing and activation-based processing. Weight-based processing in turn involves model learning (subserved by hebbian-like learning mechanisms) and task learning (subserved by error-correcting learning procedures). From the perspective developed here, activation-based processing and learning will always tend to be associated with awareness — even though it might often occur that conscious contents fail to be associated with relevant metaknowledge and therefore remain implicit. Model learning, in contrast, corresponds to the clearest case of implicit learning, to the extent that it is assumed to be a mandatory consequence of information processing. Such learning therefore never depends on the intentions or goals of the agent, and its effects, because they are very gradual, can be expressed in behavior before they become available to awareness. Task learning, by contrast, is necessarily intentional, and therefore more likely to shape representations in ways that are directly consistent with the current goals of the agent.

6. Implications

In this section, we offer a necessarily brief and sketchy set of examples where we have found our framework helpful in terms of understanding empirical phenomena such as priming, skill acquisition, automaticity, development, or the interpretation of dissociations in neuropsychology. This short review also gives us the opportunity to link our framework with similar previous accounts of these phenomena and to further contrast our own proposal with other positions.

6.1. Priming.

In a recent paper, Becker et al. (1997) describe an attractor, neural-network model of both short- and long-term priming effects that accounts for a large variety of priming phenomena as the result of an automatic process of incremental learning that is based on the same information processing and representational principles that we have just outlined. Becker and colleagues showed that semantic priming can be construed as the automatic deepening of the basin of attraction "of the semantic space for both primes and related targets, and that this effect should primarily manifest itself on semantic-retrieval tasks" (p. 1062). Their model accurately predicts that performing a semantic task on a target is influenced by having previously performed a similar task on a semantically related prime, even if a number of intervening words are presented between prime and target. Importantly, it also predicts low or null priming effects when long-term semantic effects are tested through a lexical decision task (see also Joordens & Becker, 1997) or when the processing task performed with the primes is not semantic (Friedrich, Henik, & Tzelgov, 1991; Kaye & Brown, 1985; Smith, Theodor, & Franklin, 1983). These results, as well as the successful simulation work, are compatible with our own assumptions in that they suggest (1) that learning is assumed to be a mandatory consequence of processing, (2) that the effects of learning are particularly focused on those representational features that are relevant to the processing task (and which therefore produce specific experiences); and (3) that these effects are not limited to their most direct consequences —in this case, the episodic recollection that a prime has been presented— but may also produce a host of indirect (priming) effects that are not necessarily mediated by conscious recollection of its cause.

6.2. Implicit learning

If priming can be cast as a form of implicit learning, as Becker et al. (1997) suggest, it seems that implicit learning can likewise be depicted as a form of complex relational priming. Indeed, while our framework emphasizes that learning results from conscious experience, it also makes it clear that the effects of learning need not be limited to modifying conscious experience. In particular, two important assumptions embodied in our framework are (1) that adaptation occurs as a mandatory consequence of processing, and (2) that learning has both direct and indirect effects. Three consequences of these assumptions are (1) that learning can occur without intention to learn, (2) that the changes resulting from learning can remain unconscious at the time of learning, and (3) that such changes can influence subsequent processing even in the absence of awareness that this is so.

Because of the intricate methodological issues involved, it has proved rather difficult to gather supporting evidence for any of these three claims. It is always difficult to assess exactly what participants in an experiment involving learning are actually intending to do. Implicit learning studies have often tried to circumvent this problem by exposing participants to very complex settings in which learning would not be expected to improve through an intentional orientation, but this strategy has not been frequently used (see Jiménez, Méndez, and Cleeremans, 1996, for one example). However, indirect evidence that people can effectively learn without intending to do so has been obtained through some recent experiments that use dual-cue paradigms, in which the existence of a perfect and explicit predictor of the relevant stimulus dimension can be taken to prevent the deliberate search for more complex contingencies (Cleeremans, 1997; Jiménez & Méndez, in press). The fact that learning of these complex contingencies can be obtained even under these dual-cue conditions provides us with a clear indication that this learning proceeds regardless of participants' intention to learn. Importantly, however, these results should not be taken as indicating that learning is completely unselective. Indeed, several recent experimental results (Jiménez & Méndez, 1999; 2001; see also Jiang & Chun, in press) convincingly indicate that learning is selectively obtained for those particular features that are relevant to the task(s) at hand and, hence, that learning is deeply modulated by the attentional variables that ultimately determine the learner's experiences.

As we have repeatedly stated, the fact that learning depends on the conscious experiences of the learner does not necessarily entail that all learning should be conscious at the moment of learning, or that they should be conscious to produce any effect on performance. The unconscious nature of the knowledge acquired during training on a sequence learning task has been examined by us in previous studies (e.g., Jiménez et al., 1996) and it has been recently investigated by Destrebecqz and Cleeremans (in press) by adapting Jacoby’s process dissociation procedure.

In a typical sequence learning situation (see Clegg, DiGirolamo, & Keele, 1998), participants are asked to react to each element of a sequentially structured visual sequence of events in the context of a serial reaction time task. On each trial, subjects see a stimulus that appears at one of several locations on a computer screen and are asked to press as fast and as accurately as possible on the key corresponding to its current location. Unknown to them, the sequence of successive locations follows a repeating pattern (Nissen & Bullemer, 1987), and participants learn this pattern, as showed by a progressive decrease of their reaction times, that increase dramatically when the sequential structure of the material is modified (Cohen, Ivry, & Keele, 1990; Curran & Keele, 1993; Reed & Johnson, 1994).

This learning, however, often fails to be expressed through verbal reports, (Willingham, Nissen, & Bullemer, 1989; Curran & Keele, 1993) — a dissociation that has led many authors to consider learning in this situation to be implicit. However, many of the relevant studies have been criticized on methodological grounds that would be too long to review in this chapter (but see Cleeremans, Destrebecqz, & Boyer, 1998 for a detailed overview). Suffice it to say that many of the relevant methodological difficulties stem from the fact that most empirical paradigms through which implicit learning has been studies have assumed that one take specific tasks with either implicit or explicit processing.

To overcome these methodological difficulties, Destrebecqz & Cleeremans (in press) sought to adapt Jacoby’s process dissociation procedure (e.g., Jacoby, 1991) to the study of sequence learning. Subjects were first trained, under the incidental learning conditions typical of implicit learning studies, on a second-order conditional sequence. This training occurred under two conditions defined by the length of the response-to-stimulus interval (RSI): One group of participants was trained with a standard RSI of 250 msec, and another was trained with an RSI of 0 msec. For this latter group, the next stimulus therefore appeared on the screen as soon as the previous one had been responded to. Consistently with the ideas embodied in assumption C3 above, we hoped that reducing the time available for processing would selectively impair the development of strong, explicit representations of the links between the temporal context set by previous elements of the sequence and the location of the next stimulus.

To find out about participants’ explicit knowledge of the material, Destrebecqz and Cleeremans asked them to perform two generation tasks and a recognition task. The generation task was adapted from Jacoby’s PDP, and consisted of both an inclusion task as well as an exclusion task. In inclusion, participants had to generate a sequence of 96 elements that resembled the training sequence. They were told to base their sequence either on conscious recollection or to guess. Both conscious and unconscious processes can therefore contribute to performance in inclusion. In exclusion, participants were again told to generate a sequence of 96 elements, but this time they were told to produce a sequence that was as different as possible from the training sequence. By assumption, the only way participants can perform this exclusion task successfully is by recollecting the location of the next stimulus and by selecting another location. Failure to exclude can thus be interpreted as reflecting the influence of implicit knowledge. In this condition thus, and in contrast to what happens in inclusion, conscious and unconscious components of performance act against each other. Finally, in recognition, participants were presented with 24 sequences of three elements, only 12 of which had actually been part of the training sequence. For each, they had to indicate the extent to which they believed it was part of the training sequence on a 6-points scale.

The results indicated that while both groups of participants exhibited some explicit knowledge of the material through the inclusion task, only people trained with an RSI of 250 msec were able to perform successfully in the exclusion task. People trained with an RSI of 0 msec indeed continued to generate material from the training sequence in spite of instructions to the contrary. Further, only participants trained with a 250 msec RSI were able to perform above chance on the final recognition task.

When applied to these data, our framework suggests the following interpretation: People trained with an RSI were given more opportunities to develop and link together high quality memory traces than people in the no RSI condition. Because awareness depends in part on the quality of stored memory traces, the former will therefore tend to acquire more explicit knowledge than the latter. Importantly, “no RSI” participants do acquire relevant knowledge about the sequence — but in the form of weaker memory traces that are only capable of influencing responses when contextual information is simultaneously available. This knowledge can thus be expressed in the SRT task as well as in the generation tasks because in both cases, responses can be determined based jointly on an external stimulus (self-generated in the case of the generation tasks, or produced by the experimental software in the SRT task) and the relevant memory traces. Because these traces are weak and because controlled processing (and hence awareness) requires high-quality traces to be available, their influence on performance remains undetected and controlled responding made difficult. The relevant sequential knowledge therefore cannot be inhibited when the generation task is performed under exclusion conditions. Similarly, during recognition, weak memory traces do not allow successful discrimination between old and novel sequences in the absence of perceptual and motor fluency, as was the case in Destrebecqz & Cleeremans’s study.

6.3 Skill acquisition and automaticity

Skill acquisition refers to extended periods of exposure to a particular domain during which learning occurs. It might involve learning how to use musical instrument, learning to master a particular athletic skill, or learning natural language. In our framework, skill acquisition thus involves a graded continuum expressed in terms of the relative strength of the underlying representations. This continuum involves weak, implicit representations when learning starts, and very strong, high-quality representations when training ends.

Automaticity has often been associated with lack of availability to conscious experience, but some authors (i.e., Tzelgov, 1997) have proposed that the defining feature of automatic behavior should simply be its ballistic properties, that is, the fact that once initiated, execution of the behavior can no longer be inhibited until completed. We very much agree with this position: In our framework, the strong representations associated with automatic behavior are available to subjective experience and form what one could call, along with Mangan, the "fringe" of consciousness. In other words, such representations can become conscious if appropriate attention is directed towards their contents — as in cases where normally automatic behavior (such as walking) suddendly becomes conscious because the normal unfolding of the behavior has been interrupted (e.g., because I've stumbled upon something) — but they are not normally part of the central focus of awareness nor do they require conscious control. This is reflected in our framework by assuming that the contents of phenomenal experience depend both on the potency of currently active representations as well as on their availability to control. Since availability to control is inversely related to potency for automatic representations, this indeed predicts weaker availability to phenomenal experience of very strong representations as compared to merely strong representations.

Our framework also predicts that very strong representations are left in place; that is, they become active whenever their preferred stimulus is present. This suggests that what happens over the course of learning a skill is that additional novel ways of inhibiting or otherwise modulating the effects of these very strong representations are found through processes of learning. Consider what happens when you learn to play the piano, for instance. As Karmiloff-Smith (1992) points out, one goes from effortful programming of every movement to a stage where entire sequences of movements can be executed all at once, and then to a later stage where genuine flexibility is achieved. Our suggestion here is that subjective experience at each stage simply reflects the contents of the processing modules that currently contain the most abstract representation of the stimulus. Ability to control the influence of the contents of lower-level modules is thus progressively lost during skill acquisition, but importantly, these contents are still constitutive of subjective experience, if only through their role in supporting higher-level representations.

6.4 Development

The notion that development involves continuous changes in the strength or quality of underlying representations is central in many accounts of various relevant phenomena. For instance, McClelland and Jenkins (McClelland & Jenkins, 1991)'s connectionist model of developmental changes in the balance beam task is rooted in the idea that experience at solving balance beam problems results in the progressive differential *strengthening* of the internal representations of the weight and distance information. The relatively systematic sequence of stages observed through development in the emergence of mastery on this task, which exhibits various patterns of ability to solve specific problems, can thus simply be accounted for by competition between the information-processing pathways corresponding to each dimension. In other words, this account of the emergence of skilled behavior at mastering balance-beam tasks is entirely strength-based.

Munakata and collaborators (Munakata, et al., 1997) have likewise proposed a novel account of the development of object permanence during infancy (see also Mareschal, Plunkett, & Harris, 1999) in which the notion of strength of representation also plays a central role. Classical theories of object permanence assume that at some early point during development, children “acquire the concept” that objects continue to exist when out of view. The crucial point here is that this knowledge is assumed to be of a conceptual nature: Children are taken to be constantly developing explicit theories about their environment, and their theories can be described as consisting of a set of all-or-none beliefs about the way the world works. In stark constrast, Munakata et al. suggested that the progressive emergence of appropriate anticipatory responses in situations where a moving object temporarily disappears behind a screen can emerge simply out of the operation of prediction-driven mechanisms such as insantiated in the Simple Recurrent Network (Elman, 1990; see also Cleeremans, Servan-Schreiber, & McClelland, 1989). Most importantly, Munakata et al. showed how the model, when trained on such a prediction task, progressively develops stronger, higher-quality representations of the object while it is hidden, and how this progressive strengthening of the model’s internal representations can be related to the development of knowledge about object permanence. Another important aspect of this work was the demonstration that the very same principles — strength of internal representation — could account for observed dissociations between different measures (e.g., looking times vs. reaching behavior) of children’s ability to exhibit knowledge of object permanence. It is interesting to note that in many ways, the debates elicited in the developmental literature by the empirical findings related to object permanence mirror those taking place in the field of implicit learning about whether or not subjects are best described as “knowing the rules of the grammar”.

A central idea that both of these models illustrate is that continuous changes along one dimension can exert non-linear effects on the overall behavior of the system when interactions between several dimensions are considered. In other words, all-or-none behavior can be rooted in continuous, graded changes in some relevant underlying dimension.

Finally, with respect to the development of explicit, conscious representations in cognitive systems, our framework can also be linked in interesting ways with the processes of representational redescription envisioned by Karmiloff-Smith (Karmiloff-Smith, 1992) as the main process of change during development. A crucial claim embodied in the assumptions that underpin the notion of representational redescription is that learning is success-driven, that is, behavioral mastery of a particular skill does not constitute a signal for learning to stop but rather a signal for further learning to occur — on the internal representations through which mastery was achieved. Representational description, according to Karmiloff-Smith, is a "… process of 'appropriating' stable states to extract the information they contain, which can then be used more flexibly for other purposes" (p. 25). Thus, representations change over development in such a manner that previously implicit dimensions of the problem — which are sufficient to achieve behavioral success — progressively become explicit and hence available for global control of action and for verbal report. Finally, our framework is also congruent with the idea that modules, in general, are a product of learning and development rather than their starting point.

7. Disscussion: What we leave behind

In this chapter, we have attempted to outline a framework that offers a clear functional role to consciousness by linking conscious awareness with adaptation in general, and with learning in particular. We have argued that if we take consciousness as the only mechanism through which flexible control can be achieved over action, then it follows that learning should be the most important factor that determines the contents of conscious experience. Learning thus shapes consciousness, and consciousness, in turn, reflects the adapted appreciation of the dynamics of the current situation that is necessary to make flexible control over action possible (see also Perruchet & Vinter, this volume). Our framework as it stands, however, does not address how the contents of consciousness are shaped by experience; it merely suggests the conditions under which representations are most likely to become part of conscious experience, and, importantly for our purposes, it also roots the emergence of conscious awareness into thoroughly subsymbolic mechanisms.

Further, our framework does not assume that there exists a strong distinction between conscious and non-conscious aspects of cognition. Rather, it assumes that conscious and unconscious aspects of cognition are simply that — aspects of a single set of underlying neural mechanisms. Again, this position does not deny — far from it — that there are qualitative differences between conscious and unconscious computations, but simply emphasizes that such differences are rooted in the non-linear properties of otherwise graded, continuous representation and processing systems. The most important implication of these assumptions in the context of implicit learning research is that our framework leaves open the possibility for change to occur without intention and without concurrent awareness that change is taking place.

"What we leave behind", then, is a large set of unanswered questions about the fate of what we have called "explicit representations" — those representations that we assume constitute the best candidates to form the contents of phenomenal experience. However, we hope to have convinced readers (1) that understanding conscious (symbolic) cognition necessarily involves rooting this understanding in an analysis of implicit (subsymbolic) cognition, and (2) that understanding processes of learning is fundamental for any theory of consciousness. In this respect, the study of implicit learning has a bright future, for it is through the development of sensitive paradigms through which to explore the differences between conscious and unconscious cognition that one can best contribute to the search for the neural correlates of consciousness.

Footnotes

[1] This does not necessarily imply that artificial consciousness is not possible, but simply that the relevant processes cannot consist simply of symbol manipulation.

Acknowledgements

Axel Cleeremans is a research associate with the National Fund for Scientific Research (Belgium). This work was supported by a grant from the Université Libre de Bruxelles in support of IUAP program P4/19, by grant HPRN-CT-2000-00065 from the European Commission, and by DGES Grant PB97-0525 from the Ministerio de Educación y Cultura (Spain) to Luis Jiménez.

References

Ader, R., & Cohen, N. (1985). CNS-immune system interactions: Conditioning Phenomena. *Behavioral and Brain Sciences, 8*, 379-394.

Anderson, J. R., & Memory, L. a. (1995). Learning and Memory. New York: Wiley.

Atkinson, A., Thomas, M., & Cleeremans, A. (2000). Consciousness: Mapping the theoretical landscape. Trends in Cognitive Sciences, 4(10), 372-382.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. Scientific American, 224, 82-90.

Baars, B. (1988). A Cognitive Theory of Consciousness. Cambridge: Cambridge University Press.

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. Journal of Experimental Psychology: Learning, Memory and Cognition, 23, 1059-1082.

Block, N. (1995). On a confusion about a function of consciousness. Behavioral and Brain Sciences, 18, 227-287.

Broadbent, D. E., Fitzgerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex system. British Journal of Psychology, 77, 33-50.

Chalmers, D. J. (1996). The conscious mind: In search of a fundamental theory: Oxford University Press.

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer’s innards: A psychological and philosophical perspective on the development of thought. Mind and Language, 8, 487-519.

Cleeremans, A. (1997). Principles for implicit learning. In D. C. Berry (Ed.), How implicit is implicit learning? (pp. 195-234). Oxford: Oxford University Press.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning : News from the front. Trends in Cognitive Sciences, 2, 406-416.

Cleeremans, A., Servan-Schreiber, D., & McClelland, J. L. (1989). Finite state automata and simple recurrent networks. Neural Computation, 1, 372-381.

Clegg, B. A., DiGirolamo, G. J., & Keele, S. W. (1998). Sequence learning. Trends in Cognitive Science, 2, 275-281.

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.

Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? Nature, 375, 121-123.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. Journal of Experimental Psychology : Learning, Memory and Cognition, 19, 189-202.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. Cognition, 79, 1-37.

Dennett, D. C. (1991). Consciousness Explained. Boston, MA.: Little, Brown & Co.

Dennett, D. C. (1996). Kinds of Minds: Towards an understanding of Consciousness: Basic Books.

Destrebecqz, A., & Cleeremans, A. (in press). Can sequence learning be implicit? New evidence with the Process Dissociation Procedure. Psychonomic Bulletin & Review.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. Behavioral and Brain Sciences, 22, 735-808.

Dienes, Z. & Perner, J. (this volume). A theory of the implicit nature of implicit learning.

Dretske, F. (1988). Explaining behavior. Cambridge, MA: MIT Press.

Dulany, D. E. (1997). Consciousness in the explicit (deliberative) and implicit (evocative). In J. D. Cohen & J. W. Schooler (Eds.) Scientific approaches to the study of consciousness. (pp. 179-212). Mahwah, NJ.: Lawrence Erlbaum Associates.

Elbert, T., Pantey, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. Science, 270, 305-307.

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

Eriksson, P. S., Perfilieva, E., Björk-Eriksson, T., Alborn, A.-M., Nordborg, C., Peterson, D. A., & Gage, F. H. (1998). Neurogenesis in the adult hippocampus. Nature Medicine, 4(11), 1313-1317.

Farah, M. J. (1994). Visual perception and visual awareness after brain damage: A tutorial overview. In C. Umiltà & M. Moscovitch (Eds.), Attention and Performance XV: Conscious and nonconscious information processing (pp. 37-76). Cmabridge, MA: MIT Press.

Farah, M. J. (1998). Why does the somatosensory homonculus have hands next to face and feet next to genitals: A hypothesis. Neural Computation, 10(8), 1983-1985.

French, R. M. (2000). The Turing test: The first 50 years. Trends in Cognitive Sciences, 4(3), 115-122.

Friedrich, F. J., Henik, A., & Tzelgov, J. (1991). Automatic processes in lexical access and spreading activation. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 792-806.

Frith, C., Perry, R., & Lumer, E. (1999). The neural correlates of conscious experience : An experimental framework. Trends in Cognitive Sciences, 3, 105-114.

Fodor, J. (1975). The Language of Thought. New York, NY: Harper & Row Publishers Inc.

Grossberg, S. (1999). The link between brain learning, attention, and consciousness. Consciousness and Cognition, 8, 1-44.

Holender, D. (1986). Semantic activation without conscious activation in dichotic listening, parafoveal vision, and visual masking : A survey and appraisal. Behavioral and Brain Sciences, 9, 1-23.

Holender, D. (1992). Expectancy effects, congruity effects, and the interpretation of response latency measurement. In J. Alégria, D. Holender, J. J. d. Morais, & M. Radeau (Eds.), Analytic approaches to human cognition (pp. 351-375). Amsterdam: Elsevier.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. Journal of Memory and Language, 30, 513-541.

Jiang, Y., & Chun, M. (in press). Selective attention modulates implicit learning. Quarterly Journal of Experimental Psychology.

Jiménez, L., Méndez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22, 948-969.

Jiménez, L., & Cleeremans, A. (1999). Fishing with the wrong nets: How the implicit slips through the representational theory of mind. Behavioral and Brain Sciences, 22, 771.

Jiménez, L., & Méndez, C. (1999). Which attention is needed for implicit sequence learning? Journal of Experimental Psychology: Learning, Memory, and Cognition, 25, 236-259.

Jiménez, L. & Méndez, C. (in press). Implicit sequence learning with competing explicit cues. Quarterly Journal of Experimental Psychology (A).

Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1083-1105.

Kanwisher, N. (2001). Neural events and perceptual awareness. Cognition, 79, 89-113.

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

Kaye, D. B., & Brown, S. W. (1985). Levels and speed of processing effects on word analysis. Memory and Cognition, 13, 425-434.

Kinsbourne, M. (1997). What qualifies a representation for a role in consciousness? In J. D. Cohen & J. W. Schooler (Eds.), Scientific Approaches to Consciousness (pp. 335-355). Mahwah, NJ: Lawrence Erlbaum Associates.

Klein, S. B. (1991). Learning: Principles and Applications: McGraw Hill.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. Proceedings of the National Academy of Sciences of the U.S.A., 10, 1073.

Mangan, B. (1993). Taking phenomenology seriously: The "fringe" and its implication for cognitive research. Consciousness and Cognition, 2, 89-108.

Mareschal, D., Plunkett, K., & Harris, P. (1999). A computational and neuropsychological account of object-directed behaviours in infancy. Developmental Science, 2, 306-317.

Mathis, W. D., & Mozer, M. C. (1996). Conscious and unconscious perception: A computational theory. Paper presented at the Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society, Hillsdale, N.J.

McClelland, J. L. (1979). On the time-relations of mental processes : An examination of systems in cascade. Psychological Review, 86, 287-330.

McClelland, J. L., & Jenkins, E. (1991). Nature, nurture, and connectionism: Implications for connectionist models of development. In K. v. Lehn (Ed.), Architectures for Intelligence — The Twenty-second (1988) Carnegie Symposium on Cognition . Hillsdale, NJ: Lawrence Erlbaum Associates.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychological Review, 102, 419-457.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. Psychological Review, 10(4), 686-713.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirement of learning: Evidence from performance measures. Cognitive Psychology, 19, 1-32.

O'Brien, G., & Opie, J. (1999). A connectionist theory of phenomenal experience. Behavioral and Brain Sciences, 22, 175-196.

O'Reilly, R. C., & Munakata, Y. (2000). Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain. Cambridge, MA.: MIT Press.

Perruchet, P., & Vinter, A. (1998). Learning and development: The implicit knowledge assumption reconsidered. In M. A. Stadler & P. A. Frensch (Eds.), Handbook of implicit learning (Vol. 15, pp. 495-531): Sage Publications.

Perruchet, P., Vinter, A., & Gallego, J. (1997). Implicit learning shapes new conscious percepts and representations. Psychonomic Bulletin and Review, 4, 43-48.

Perruchet, P. & Vinter, A. (this volume). The self-organizing consciousness: A framework for implicit learning.

Posner, M. I., & Rothbart, M. K. (1998). Attention, self-regulation, and consciousness. Philosophical Transactions of the Royal Society B, 353, 1915-1927.

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. Journal of Experimental Psychology: Learning, Memory and Cognition, 20, 585-594.

Rosenthal, D. (1986). Two concepts of consciousness. Philosophical Studies, 94, 329-359.

Rosenthal, D. (1997). A theory of consciousness. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), The Nature of Consciousness: Philosophical Debates. Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations. Cambridge, MA: MIT Press.

Schacter, D. L. (1989). On the relations between memory and consciousness: Dissociable interactions and conscious experience. In H. L. Roediger and F. I. M. Craik (Eds.), Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving (pp. 355-389). Mahwah, NJ: Lawrence Erlbaum Associates.

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3, 417-457.

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

Searle, J.R. (1999). Chinese Room Argument. In R.A. Wilson and F.C. Keil (Eds.) The MIT Encyclopedia of the Cognitive Sciences (pp.115-116). Cambridge, MA: The MIT press.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. Behavioral and Brain Sciences, 17, 367-447.

Smith, M. C., Theodor, L., & Franklin, P. E. (1983). The relationship between contextual facilitation and depth of processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 697-712.

Tarpy, R. M. (1997). Contemporary Learning Theory and Research: McGraw Hill.

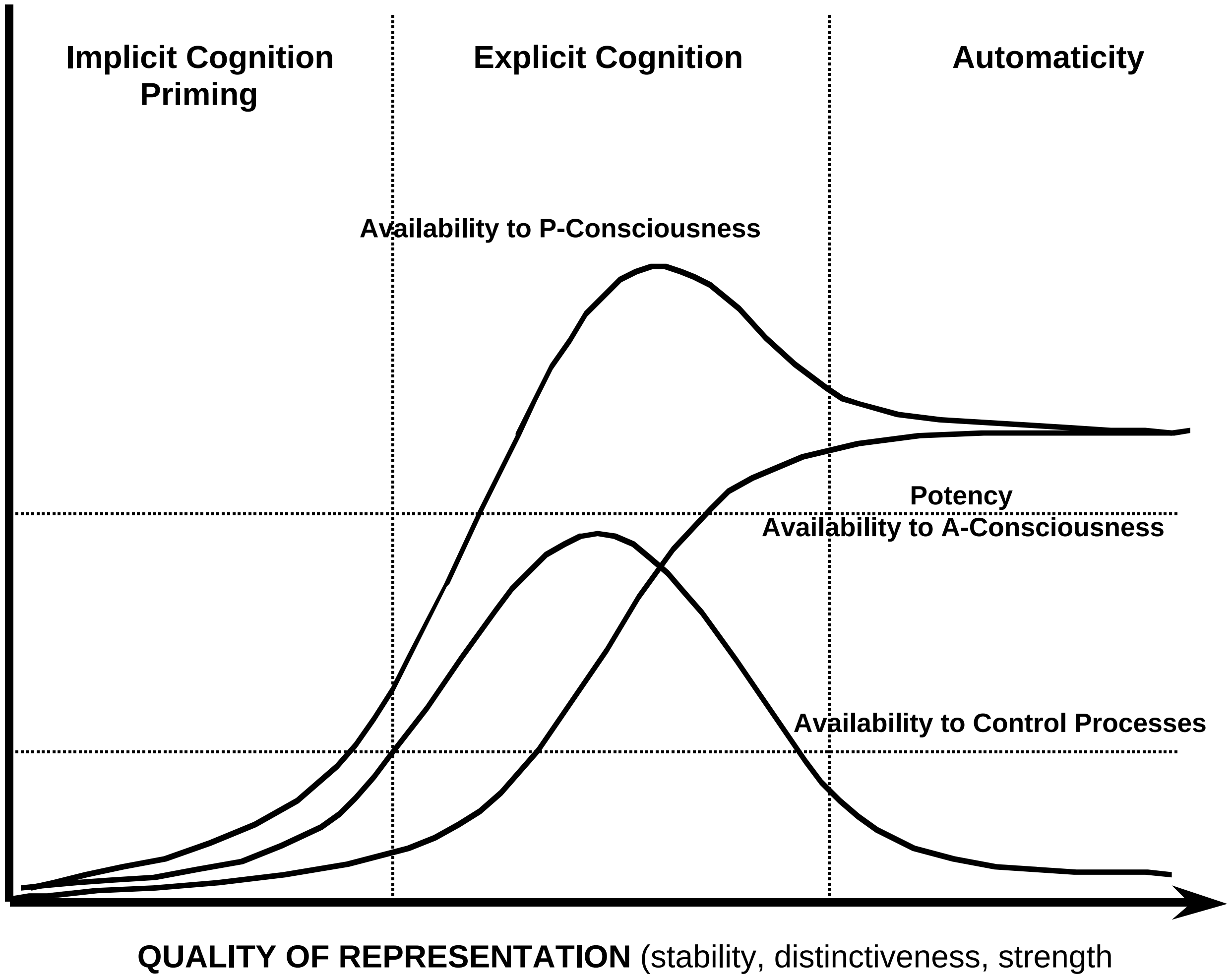
Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. Science, 282(5395), 1846-1851.

Tzelgov, J. (1997). Automatic but conscious: That is how we act most of the time. In R. S. Wyer (Ed.), The automaticity of everyday life (Vol. X, pp. 217-230). Mahwah, .N.J.: Lawrence Erlbaum Associates.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. Journal of Experimental Psychology : Learning, Memory and Cognition, 15, 1047-1060.

Figure Captions

Figure 1: Graphical representation of the relationships between quality of representation (X-axis) and (1) potency, (2) availability to control, (3) availability to subjective experience. See text for further details.


Implicit Cognition
Priming
Explicit Cognition
Automaticity
Availability to P-Consciousness
Potency
Availability to A-Consciousness
Availability to Control Processes
QUALITY OF REPRESENTATION (stability, distinctiveness, strength