Introduction

Robert M. French
Quantitative Psychology & Cognitive Science
Psychology Department
University of Liège,
4000 Liège, Belgium
email: rfrench@ulg.ac.be

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

The study of consciousness spans a host of disciplines ranging from philosophy to neuroscience, from psychology to computer modeling. Arguments about consciousness run the gamut from tenuous, even ridiculous, thought experiments to the most rigorous neuroscientific experiments. This book offers a novel perspective on many fundamental issues about consciousness based on empirical, computational and philosophical research on implicit learning — learning without conscious awareness of having learned.

There are many profound and interesting issues involving consciousness that fall within the purview of serious science. Indeed, the "search for the neural correlates of consciousness" has now become a major endeavor in the cognitive neurosciences, as evidenced by countless articles, major books and scientific meetings. This spectacular renewed interest in the biological bases of consciousness can no doubt be attributed to the increasingly widespread availability of sophisticated brain imaging techniques, which, for first time, make it possible to conduct detailed explorations of the correspondences between subjective (i.e., mental) and objective (i.e., neural) states. This empirical program, however, cannot in and of itself answer all the questions one might have about consciousness. Consciousness, indeed, is a complex phenomenon that poses unique conceptual and methodological challenges. A first challenge is conceptual: How do we best characterize the various dimensions of consciousness? Consciousness is not a single thing, but includes different dimensions, such as subjective experience, intentional control and attention, or self-consciousness. Whether these different dimensions of consciousness are dissociable, and whether or not they involve the same neural mechanisms are questions that continue to elicit lively debate among philosophers, cognitive psychologists and neuroscientists alike. A second challenge is methodological: How do we best differentiate between conscious and unconscious cognition? What operational criteria should we use to decide whether a subject is conscious of some aspect of a particular situation? To what extent is unconscious cognition possible?

As it turns out, implicit learning is one of the paradigms (along with subliminal perception) in which such issues have been explored most thoroughly. Because the field as a whole has been particularly concerned with the delicate methodological issues involving in attempting to establish relationships between subjective and objective measures of performance, it is also one of the most interesting paradigms through which to start exploring consciousness itself. This book is an attempt to come to grips with some of the most significant issues surrounding implicit learning and its relationship to

conscious awareness. By bringing together a number of the most important figures in implicit learning research, we have attempted to present the major trends in the field and to show how these trends might gradually be converging towards a consensus. The methods involved in these explorations range from the philosophical to the clinical and from the experimental to the computational. As editors, we have resisted any attempt to select our authors based on any particular conceptual framework or point of view. Rather, we felt that it was important to present a broad spectrum of approaches to this fascinating area, with a specific focus on integrating contributions that address the three central issues raised by the large body of research on implicit learning — namely, (1) to what extent can learning be unconscious?, (2) what is the nature of the knowledge acquired in implicit learning situations?, and (3) to what extent should implicit and explicit learning be characterized as involving separable, independent processing systems?

The field of implicit learning can reasonably be said to begin with the publication of Arthur Reber's 1967 paper entitled "Implicit learning of artificial grammars". Reber himself begins our book with a series of reflections that not only cover the 35 years since his seminal paper appeared, but also identify what he considers to be the central issues that will set the agenda for future research in the field.

Axel Cleeremans and Luis Jiménez then attempt to develop an overall framework for implicit learning. Their central thesis is that mechanisms underlying both conscious and unconscious (i.e., explicit and implicit) learning are really the same and that it really is only a difference of degree that gives rise to one type of learning or the other. They regard the central function of conscious awareness to be an evolutionary adaptive means of providing us with flexible control of our behavior. Given the importance of consciousness to their account of implicit/explicit learning, the authors spend a considerable amount of time discussing questions related to consciousness, in general, and computational accounts of consciousness, in particular. One of their central claims is that consciousness is not an all-or-nothing process, but is rather a multi-faceted, graded phenomenon. This contrasts rather sharply with one major perspective, represented in these pages, that denies the existence of unconscious mental representations. For Cleeremans and Jiménez there is an ongoing and interactive relationship between consciousness and learning. While they take a dynamic approach to learning, they are not hardcore advocates of a "dynamical systems" approach to cognition. Rather, they believe in the existence of graded, internal representations of an external reality, very much in keeping with the philosophy of the distributed neural network models that have been on the scene for the past two decades.

Cleeremans & Jiménez spell out the interaction between learning and consciousness and claim, significantly, that the extent to which a particular representation is available to consciousness depends on its quality (defined in terms of activation level, stability in time, and contextual distinctiveness), that learning gradually produces better adapted representations, and that the function of consciousness is to control those representations that are able to influence behavior. They appeal to a "representation-quality" model of conscious awareness to make their case for implicit learning, to wit: If the quality of a learned representation is good enough, it is perceived consciously; if not it will remain

unconscious (and thus learning, in this case, will have been implicit). They describe an elegant experiment in which a technique called the Process Dissociation Procedure is used to show that learning remains largely implicit as long as the cognitive system is not given enough time to develop high quality representations. In summary, this chapter brings a large body of theoretical as well as empirical evidence to bear on their central claim that conscious and unconscious awareness — and the associated types of learning — merely represent different points on the same continuum of underlying neural mechanisms.

The pendulum swings back from a Cleeremans and Jiménez's graded, dynamic stance to Pierre Perruchet and Annie Vinter's denial of unconscious mental representations. Their view is unambiguous in the extreme: "Processes and mechanisms responsible for the elaboration of knowledge are intrinsically unconscious, and the resulting mental representations and knowledge are intrinsically conscious. No other components are needed." In other words, at the heart of their position is the explicit refutation of the notion of unconscious representations. Perruchet & Vinter point to an overwhelming body of evidence that shows that attention to the initial sensory data is necessary for improved performance on implicit learning tests. But this would seem to lead to what appears to be a major paradox: Why should the initial coding of the sensory data require conscious attention, while all the subsequent operations performed on those data require none?

They provide a two-fold answer to this question via the notion of self-organization of perceptual experience and the results of a computational model, PARSER. They begin by showing how chunking and segmentation in an undifferentiated stream of utterances gradually emerges both in humans and in their model. One of their key claims is that the discovery of words results from the fact that "the probability of repeatedly selecting the same group of syllables by chance is higher if these syllables form intra-word rather than between-word components." They extend this principle of word segmentation to world segmentation and arrive, perhaps unwittingly, at a re-statement of the underlying theoretical principle for the emergence of Roschian basic level categories — namely, that a basic level category is a category grouping where the ratio of inter-category variance to intra-category variance is maximized. Perruchet and Vinter suggest that this unconscious process of self-organization of perceptual experience gives rise to conscious representations of sensory input. In short, implicit learning is better conceived of as a transformation of conscious experience through the action of elementary associative learning and memory processes acting on components of these experiences. In this respect at least, there is significant agreement between this framework and that of Cleeremans and Jiménez: Both proposals indeed assume that implicit learning shapes conscious experience.

The next chapter, authored by Zoltán Dienes and Josef Perner, returns the discussion to a more foundational level. Dienes & Perner are essentially concerned with developing a carefully crafted definitional framework for explicit and implicit learning, providing a detailed exploration of the relationships between representation, consciousness, metacognition and, of course, implicit/explicit learning. On the matter of representation,

they adopt a functionalist stance: representations represent something because of the functional role they play in the cognitive economy. In this they follow Fred Dretske and claim that "A represents B just in case A has the function of indicating B." Dienes & Perner maintain that this definition of representation is broad enough to encompass what cognitive scientists generally mean by representation. (Later in their chapter they tackle the harder question of what makes certain representations conscious and others not.) Based on this definition, they delineate explicit from implicit representations as follows: "Any environmental feature or state of affairs that is not explicitly represented but forms part of the representational content is represented implicitly." The idea here is that when learning a piece of knowledge, K, any information that is a necessary supporting fact of K, but is not present explicitly in K, is implicit knowledge. Thus, implicit in the fact that "Bill is a bachelor," is the necessary supporting fact that Bill is male, even though this is not explicitly represented in the original statement.

The authors then go on to clarify the notion of explicit representation and suggest three levels of explicitness, tying them to consciousness by means of second-order knowing, i.e., knowing that we know. For them, a second-order thought is always necessary for conscious awareness of an event and fully explicit knowledge is necessary for any knowledge to be conscious knowledge. They show how the distinction between declarative knowledge and procedural knowledge fits into this framework. Next, two types of voluntary control are introduced and related to metacognition, the essence of which, they believe, consists of monitoring and control. This then gives them a yardstick by which to judge implicitness: implicit processes essentially lack various degrees of metacognition, while explicit processes centrally involve metacognition. Finally, the authors define implicit learning within the framework they have built up: Implicit learning is a type of learning that results in knowledge "which is not labeled as knowledge by the act of learning itself," whereas explicit learning results in knowledge that is so labeled. In short, explicit learning produces conscious, fully explicit knowledge; implicit learning produces knowledge of which we are unaware. Importantly, Dienes and Perner bring a wealth of experimental data to buttress their claims and distinctions. Unlike Perruchet & Vinter, they firmly believe that implicit representations, even though they do not constitute conscious knowledge, can control action.

In the next chapter David Shanks, Theresa Johnstone and Annette Kinder introduce an "episodic-processing account" of implicit learning. This chapter amounts to a frontal assault on dual-system accounts of learning that assume the existence of separate learning systems for implicit (general, abstract, procedural) knowledge and explicit (episodic, specific, declarative) knowledge. They begin by laying out four major claims of proponents of dual-system theories. Their episodic-processing account is then compared to rule-abstraction accounts of learning, and areas in which these accounts disagree are pinpointed. Specifically, these differences involve whether we process episodes or use rules, as well as the degree to which the test instructions elicit an implicit or explicit expression of the knowledge.

They consider possible forms of knowledge that could be used in classification and recognition tasks, starting with abstract-rule knowledge and evidence for it from cross-

modality transfer results. The authors reject these results, claiming that ultimately there is little, if any, evidence for cross-modality transfer. They move on to strict exemplar-based accounts of knowledge in which classification of new items is achieved based on their similarity in memory to specific stored training examples. Next, they examine the notion that participants learn about the frequency of occurrence of *fragments* (two- and three-letter strings) in the training strings and classify new strings based on these fragments. Finally, evidence that classification is done by a combination of rules and fragments of knowledge is briefly reviewed.

The introduction of Shanks and colleagues' episodic-processing framework follows. This account suggests (1) that processing knowledge is acquired in addition to structural knowledge of the training stimuli; (2) that training instructions significantly influence how the training items are encoded and (3) that the same knowledge can be used either implicitly or explicitly depending how this knowledge was acquired. In particular, on this account, participants learn the particular aspects of the stimuli that are relevant to the task they are engaged in, which provides affordances for subsequent, related tasks.

Finally, the authors consider the case for the existence of two distinct learning systems, based on a particular dissociation in amnesics — namely, that it has been shown repeatedly that, while amnesics' declarative memory may be poor, their non-declarative memory frequently remains largely intact. The authors point out a number of problems with Knowlton & Squire's well known 1993 study of amnesics that led them to argue for a dual-memory system. Shanks and his colleagues then point to two single-system models capable of reproducing the dissociations that Knowlton & Squire observed. The first single-system model they discuss is Nosofsky's Generalized Context Model, followed by their own model, an Elman network, that also shows this dissociation. Their model shows excellent fit to data and they conclude, reasonably, that a dual-memory system is not necessary to account for the amnesic dissociation data.

Martin Redington and Nick Chater consider the notion of knowledge representation and transfer in their chapter. They begin by reviewing what is normally meant by "transfer" in the artificial grammar learning (hereafter, AGL) paradigm. They continue by suggesting that transfer effects, which they acknowledge exist, are construed by most of the AGL community to demonstrate that the knowledge learned is represented in terms of rules encoded in a "surface-independent" format (i.e., not in its original form, but rather in an abstract form). The authors take issue with this claim, and argue instead that surface-independence and rule-based knowledge are orthogonal concepts. Redington and Chater go on to identify three distinct kinds of representation in the AGL literature — namely, knowledge of whole exemplars, knowledge of fragments of the training items, and rule-based knowledge. Their claim is that all three kinds of knowledge can, in principle, be tied either to a particular kind of surface encoding or, alternately, can be encoded in a surface-independent manner. In other words, the manner of encoding is independent from the type of knowledge encoded.

Transfer and surface-independent encoding is then discussed at length, accompanied by various accounts of transfer, for each of which they show that the knowledge acquired is

bound to the *original* surface form of the training items. They tie this to the broader issue of knowledge representation in AGL and re-iterate their claim that the mere existence of transfer does not necessarily imply that the knowledge learned is encoded in a surface-independent manner. They point to both empirical as well as computational studies that emphasize the importance, possibly the necessity, of surface-based knowledge encoding. The authors then discuss in detail an early experiment by Arthur Reber that purported to show strong evidence for surface-independent knowledge. However, the authors were unable to replicate Reber's results in two separate experiments and they conclude, modestly, that their results leave Reber's findings "open to question." In summary, their results show that "once the surface form of the materials is changed, any memorization advantage for previous exposure to the grammar disappears," which strongly argues against surface-independent encoding of knowledge under normal implicit learning conditions.

The authors then review the question of surface-independent knowledge acquisition being dependent on the context of learning and the age of acquisition of the knowledge. They accept that under certain specific conditions, surface-independent knowledge acquisition might be possible. They conclude with a discussion of "lazy" and "eager" learning in which the learner stores information, in the former case, largely in unprocessed form and, in the latter, actively attempts to extract regularities from new items. According to the authors, it is likely that we use some mixture of these two learning processes when acquiring new information from our environment. The argue that if these processes are active for learning, in general, they are likely active for implicit learning in particular. They conclude this chapter with the suggestion that, possibly, the ability to find abstract regularities (i.e., surface-independent encoding) could obtain in adults using more natural, speech-like materials.

Finally, the book concludes with an empirically oriented, clinical chapter by Thierry Meulemans and Martial Van der Linden. The main purpose of their chapter is to present data obtained from amnesic patients that bear on the debate about the implicit vs. explicit nature of knowledge acquired in implicit learning tasks. Specifically, these patients are dramatically impaired for explicit (episodic) memory tasks, while their performance on implicit memory tasks, such as AGL and serial reaction time tasks, is largely preserved. In addition to their review of the literature on implicit learning in amnesia, the authors also present a study in which they explore the implicit learning abilities of amnesic patients by using an AGL task in which the test strings were constructed in such a way that grammaticality judgments could not be based on superficial features of the learning strings. They also investigate the validity of an explicit sequence generation-task is order to assess the explicit knowledge acquired in an artificial grammar learning task. Using the AGL paradigm, the authors show that amnesic patients and controls performed at the same level during the classification task, whereas amnesic patients performed worse than controls on the generation task. Their results also showed that performance in the generation task is directly related to information learned during the study phase and not to information presented during the classification phase. Moreover, there was no correlation between the implicit and explicit measures. These results are compatible with the hypothesis of the existence of two different kinds of representation in artificial grammar

learning: the first based on processes involving fragment-specific knowledge (the chunks, which can be accessed explicitly), the second based on the learning of simple associations and more complex conditional relations between elements. Patients' performance on the classification task depends primarily on this latter mechanism, which seems to be preserved in amnesia and which can therefore be considered as being implicit.

In conclusion, we hope to have convinced our readers of the importance of a broad, multidisciplinary approach to the study of implicit learning and consciousness. We believe that new tools, in particular, neural imaging and sophisticated neural network models implemented on ever more powerful computers, will bring us closer to a consensus on precisely what is meant by implicit learning, on how to best measure it, and on its relation to consciousness. This book represents a small step in that direction.

Acknowledments

The editors would especially like to thank Tim Valentine and Vivien Ward for their contribution to the negotiations that made this book possible in the first place. Thanks to Rachel Brazil for actually getting the book into production. We would also like to acknowledge the support of the University of Liège, the Belgian Fund for Scientific Research (FNRS), and a grant from the Belgian government, IAUP 4-19, that made possible the 1998 Implicit Memory and Learning Workshop at the University of Liège, where the idea to bring out this book originated. Finally, we would like to acknowledge the generous support of the European Commission through grant HPRN-CT-2000-00065 to Robert French. Axel Cleeremans is a Research Associate of the National Fund for Scientific Research (Belgium)

Robert M. French
Axel Cleeremans

Liège-Brussels, May 2001