

1 Introduction

Every extension of knowledge arises from making the conscious the unconscious – Nietzsche

Organisms need to learn. Learning provides the basis for adaptation to a diverse and changing environment. Those organisms that are better at learning are more successful both in the acquisition of resources and the avoidance of potentially detrimental situations. When learning and memory are discussed in an everyday context, it is usually to retain information for a test or to avoid forgetting a person's name. This type of memory usually involves facts or concepts and is described as explicit. There are cases where individuals have shown an extraordinary associative capacity to the point where sensory associations from their explicit memories overpower current experience.

One such individual is the famous Solomon Shereshevskii, known in the literature simply as 'S'. The extent of his success (and difficulties) was described by the neuropsychologist Alexander Luria (Luria, 1968). From the Wikipedia entry for Shereshevskii, 'S' was a Russian journalist whose abilities were discovered when he was scolded for not taking notes at a speech. When questioned, 'S' was able to recite the speech verbatim. 'S' was tested for decades and his abilities were exceptional; in a matter of minutes, he was capable of memorizing text in languages he had never been exposed to. In his book, Luria describes the abilities of 'S' as being related to synaesthesia, the experience of sensation in one sensory modality when presented with a stimulus in another (for example, smelling color). 'S' formed complex representations of meaningless symbols that bridged multiple modalities. A nonsense syllable might produce a sharply shaped cloud that tasted sour, for example. Refining these associative

abilities, he became a very successful mnemonist, able to retain a great deal more than he could without the techniques. Most surprising was that 'S' scored absolutely average on intelligence tests. While his abilities sound incredibly useful to anyone who has ever stood stammering while trying to recall a name or reference, this ability had its cost. The extent of his associations eventually left him unable to interact normally; for instance he once explained that he was unable to eat strawberry ice cream because the tone of the ice cream's vendor left the taste of coal in his mouth. He spent the later portion of his life in an asylum. 'S' is an example of nearly perfect explicit association; each stimulus had such a rich sensory representation that he was able to form episodic memories that were very robust.

A type of learning that is better studied and more common is that of conditioned association, often referred to as implicit since it does not require the subject to be aware of the association. Conditioning is studied in a wide variety of organisms from mollusks to fruit flies, rodents, monkeys and humans (Baer and Fuhrer, 1982; Mackintosh, 1983; Gallistel, 1990; Thompson and Krupa, 1994; Connolly et al., 1996; Eichenbaum, 1997; Pearce et al., 1997; Tully, 1998; Squire and Kandel, 1999; Kocorowski and Helmstetter, 2001). This gives scientists a large number of tools to study how this association takes place. The most notable model used to study association is that of Pavlovian conditioning.

1.1 Explicit and Implicit Aspects of Conditioning

Pavlovian conditioning involves the association of a previously neutral stimulus, such as a bell, with a meaningful stimulus, such as food. Initially, the subject of study has a

reaction, such as salivation, to only the meaningful stimulus. Over time, the subject begins to respond to the previously neutral stimulus in the same way as the meaningful one. The subject has formed an association; he or she now begins salivating to the presence of the bell alone without food. In Pavlovian conditioning terms, the initially neutral stimulus (the bell) is referred to as the Conditioned Stimulus or CS. The initially meaningful stimulus (the food) is referred to as the Unconditioned Stimulus or US.

This thesis examines the interaction between implicit and explicit learning. The simple association that takes place in most organisms is often described as *implicit*, occurring without any relationship to conscious knowledge (Manns et al., 2002). Learning a person's name or associating two abstract concepts is described as *explicit* since it occurs, essentially by definition, with conscious knowledge.

It is easy to understand how conditioning could be considered an implicit process; even the 302 neurons in the roundworm *Caenorhabditis elegans* show learning in a large number of association paradigms (Rankin et al., 1990; Wen et al., 1997; Law et al., 2004). With such a small number of neurons involved, it becomes more difficult to imagine that explicit knowledge is involved in the process.

Although the differentiation between "explicit" and "implicit" has intuitive weight, what makes a particular learned relationship explicit? The study of a patient referred to as HM served as a means to split the two learning systems. HM's case was first described by Scoville and Milner in 1957 (Scoville and Milner, 1957). HM was involved in an accident at a young age that eventually resulted in epilepsy. His condition was bad enough that the medial temporal lobes, including an area called the hippocampus, on both sides of his brain were removed. After the operation, HM was unable to form new

memories. He retained older memories, but was unable to recall what he had eaten for breakfast or what he had done yesterday. He could carry on a conversation as long as there weren't too many changes of topic, and as long as it was less than a few minutes. While HM could not form any new explicit memories, his implicit learning remained intact. Over several days he learned to trace objects in a mirror without ever being able to report that he had tried the task before. He could learn new skills, motor associations, without any explicit knowledge. This finding has been generally interpreted to mean that explicit and implicit learning depend on two separate memory systems (Squire and Kandel, 1999).

A set of experiments by Larry Squire and colleagues described a further dissociation between explicit and implicit learning using two different types of eye-blink conditioning. In eye-blink conditioning, a neutral stimulus, such as a tone (the CS), is paired with a puff of air to the eye, the US. As the experiment progresses, the subject learns to blink (non-consciously) at the appropriate time to protect their eye.

Delay conditioning is an example of Pavlovian conditioning that has been described as occurring independently of awareness (Manns et al., 2002). For the association to be described as delay conditioning, the CS must precede the US, either overlapping with it or directly before it. It is called delay conditioning because the US presented is *delayed* with respect to the CS.

Trace conditioning is an example of Pavlovian conditioning that is believed to occur only when the subject has acquired conscious knowledge (Clark and Squire, 1998). The difference between delay and trace conditioning is that in trace conditioning, the CS is separated in time from the US. It is called *trace conditioning* because the association of

the neutral (CS) and meaningful (US) stimulus requires that a *memory trace* of the CS be kept after the CS terminates in order to associate it with the US. In spite of its relative complexity, there is evidence that this type of learning occurs in the fruit fly *Drosophila melanogaster* (Tully and Quinn, 1985).

There are many examples of explicit influences in implicit processes. Studies from as early as 1937 describe conditioning physiological responses using only verbal instruction (Cook and Harris, 1937); that is, the subject begins to respond to a previously neutral stimulus because of false instructions that say the stimulus may now be paired with a shock.

1.2 Our Approach

We sought to further examine the interactions between explicit and implicit processes in a classical conditioning paradigm. We chose to use a fear conditioning paradigm rather than an eye-blink conditioning paradigm because fear conditioning is well studied in a wide variety of organisms. Fear conditioning's widespread use provided us not only with better information about the paradigm being studied, but also made model systems available where electrophysiology or lesion tools could be used. It was also clear that past work with rats could be reliably and quickly reproduced in mice, a model system our lab was interested in working with for the genetic tools available. The first section of this thesis was completed in collaboration with labs at Caltech (David Anderson and Henry Lester) and UCLA (Michael Fanselow). The aim of this collaboration was to design and perform experiments with the other labs in mind, making what was learned from studies of explicit learning in humans applicable to rodent systems, and what was learned from

lesion studies in rodents applicable to behavioral results with human subjects. Details of the collaboration are contained in the discussion. This thesis is organized in three main parts.

Chapter 2 describes the effects of performing a working memory task during delay and trace conditioning. We reasoned that if trace conditioning depended on high level mental resources, such as working memory, then having subjects perform a working memory task during conditioning would eliminate trace conditioning, leaving delay unaffected. Instead, we discovered that the working memory task affected not only trace conditioning, but delay as well. These effects could be partially overcome by simplification of the protocol; for example, reducing the number of stimuli or providing the subject with information before the experiment. This study provides strong evidence of the influence that explicit processes have on implicit ones.

Chapter 3 describes areas of the brain that are important for the acquisition of both explicit and implicit information. Subjects were aversively conditioned using both trace and delay protocols during fMRI acquisition. We also recorded skin conductance responses for use as a correlate of implicit learning, and shock expectancy responses as a correlate of explicit learning. Our analysis identified portions of the brain where hemodynamic responses correlated with both of these measures. Consistent with the result of Chapter 2, the middle frontal gyrus, an area associated with working memory performance, correlated with the accuracy of shock expectancy. We also found that the amygdala, hippocampus, and areas of visual cortex correlated with the implicit measure.

The identification of visual cortex as an area correlating with implicit learning led us to specifically examine the visual areas representing the CS.

Chapter 4 describes an experiment assessing blood oxygenation changes in order to assess differences in the fusiform face area that occur as a result of delay conditioning to images of faces. These changes are consistent with an increase in the representation of the paired stimulus. They are also persistent beyond extinction of the conditioned association, indicating lasting changes in visual cortex that could continue to have effects on explicit responses well after any aversive response is removed. This study was also designed to show that the stimuli used in Chapter 2 caused similar changes in visual cortex. There was no substantial evidence indicating that this was true. Possible reasons for this are discussed in Chapter 4.