Amnesia, Memory, and the Implicit/Explicit Learning Distinction

J. Brendan Ritchie (SENECA.Y@Gmail.Com)
Department of Philosophy, 453 University College
Winnipeg, MB, R3T 2M8 Canada

Abstract

Whether or not a disassociation between implicit and explicit learning can be proven experimentally has been widely contested. In this paper I am concerned with the arguments that David Shanks and colleagues have made for why results from work on amnesia are: (a) not relevant to the debate (Shanks & St. John, 1994), and (b) can be simulated by connectionist models that do not involve distinct learning systems (Kinder & Shanks, 2001, 2003): making amnesia irrelevant. None of the arguments that Shanks and colleagues make succeed at proving either (a) or (b).

Keywords: amnesia; implicit learning; explicit learning
episodic memory; disassociation; connectionism; hippocampus.

Background

Since Reber (1967) first used the term ‘implicit learning’ there has been much debate as to whether or not a distinction can be shown between implicit and explicit forms of learning: this being especially true of the last 15 or so years (Shanks 2005: 204).

Naturally, various methods have been used to try and prove the case, both for and against, the implicit/explicit learning distinction. Especially common, for example, are the Artificial Grammar Learning (AGL) experiments first conducted by Reber (1967). Reber had subjects study seemingly nonsense letter strings, such as VXVS, and then told them that the strings were constructed using a set of rules. The subjects were then informed that in the next stage of the experiment they would have to try and determine which strings were grammatical, and which were not. Subjects displayed an above chance ability for identifying grammatical letter strings, but could not report what the grammatical rules were. Thus, it was concluded, the rules had been learned implicitly.

Such work, however, is far from conclusive. For instance, Perruchet & Pacteau (1990) found that subjects are successful at determining which strings are grammatical if they only study letter pairs, such as VX, and not full letter strings. If this is right, then it is mistaken to expect subjects to report rules, since knowledge of pairs is enough for determining which strings are grammatical. The hypothesis being that learning letter pairs caused the desired results, which subjects can report familiarity with, rather than implicitly learned rules.

Despite such difficulties many researchers (e.g. Knowlton, Ramus & Squire, 1992) think that methods like AGL experiments can overcome such difficulties by using amnesiacs in the place of normal subjects. In light of this, whether studies of amnesiacs provide any novel defense of the implicit/explicit learning distinction has been questioned. Especially prominent in this regard has been the work of David Shanks and colleagues (Shanks & St. John 1994; Kinder & Shanks 2001, 2003).

In the remains of this paper I take up the arguments and results of Shanks and Mark St. John (1994), as well as Shanks’ work with Annette Kinder (2001, 2003). I will argue such research fails to show the irrelevance of results from studies of amnesiacs.

While not the only researchers to be critical of multi-model theories of learning and memory (e.g. see Roediger 1990) the work of Shanks and others is especially relevant. This is because in the case of Kinder & Shanks they focus solely on amnesia and implicit learning, rather than implicit forms of learning and memory, and, in the case of Shanks & St. John are widely cited, so it is important to see where their arguments fail.

The Irrelevance of Amnesia

This section is split into two parts: part one responds to the target article of Shanks & St. John (1994) who argue that evidence from work on amnesia is undermined by the fact that some amnesiacs suffer from only partial deficits to explicit learning. Part two objects to the arguments made in response to commentary on Shanks & St. John (1994).

Implicit Learning and Implicit Retrieval

According to Shanks & St. John implicit learning occurs if there is (i) learning that is not accompanied with awareness of the information being learned, and (ii) a separate system from the one that operates during more common forms of learning (1994: 368). In contrast, implicit retrieval is retrieval where the information from a past experience is recalled and influences current processing without conscious recall of the prior experience (Shanks & St. John, 1994: 372). Shanks & St. John hold that, while a case of implicit learning entails the occurrence of implicit retrieval, it is not the case that implicit retrieval entails implicit learning.

As an example, suppose a subject emits a galvanic skin response (GSR) in response to a conditioning stimulus (CS): a tone; the CS previously being paired with an unconditioned stimulus (US): a shock. When the subject emits the GSR, one of three things might have occurred:

1. The subject is able to explicitly recall the study episode: the subject can verbally report or somehow show evidence of recalling the study episode. Perhaps the subject remembers where the shock was administered, or from where the tone was being emitted. This would fail to be a case of implicit retrieval, since the subject can consciously recall the study episode.

---

1 This definition follows that of Schacter (1987).
2. The subject cannot consciously recall the study episode, but shows evidence of semantic memory of the learning episode: the subject might know that the tone precedes a shock, but not where the information was learned. According to Shanks & St. John, this would count as a case of implicit retrieval, but not as a case of implicit learning.

3. The subject cannot recall the study episode, nor can the subject show evidence of semantic memory. In contrast to the last scenario, the subject would not know that the tone precedes a shock. This would count as a case of both implicit retrieval and learning.

As stated, (2) and (3) are cases of implicit retrieval, but only the latter is a case of implicit learning (Shanks & St. John, 1994: 372). If it were the case that amnesiacs provide evidence for implicit learning, then amnesiac subjects would exhibit behaviour characteristic of (3), and according to Shanks & St. John they fail to do so. This is because on the one hand an amnesic patient will respond normally to a CS, but fail to recall the study episode. But, on the other hand, the amnesiac subject “may well have been” able to verbally report that the US is followed by the CS (Shanks & St. John, 1994: 394). The ability of amnesics to show evidence of semantic knowledge in a GSR test means they do not display behaviour found in (3). Following the episodic/semantic distinction of Tulving (1983), Shanks & St. John suggest that while many studies (e.g. Weiskrantz & Warrington, 1979) show a failure of episodic memory, there is evidence of conscious rule learning in amnesics (e.g. Wood, Ebert & Kinsbourne, 1982), which is comparable to showing less than a total failure of semantic memory. Since episodic and semantic memory are thought to be forms of explicit memory, the fact that amnesics only show a failure of one kind of explicit memory does not suggest a disassociation between the analogous distinction between explicit and implicit learning.

I will assume that there is good reason to think that some amnesiac subjects seem to suffer from a failure of episodic memory, but not of semantic memory. Vargha-Khadem, Gadian, Watkins, Connelly, Van Paesschen & Mishkin (1997) discuss three cases of anterograde amnesia where the subjects were able to acquire factual information, in spite of their severe anterograde amnesia. In all three cases, the damage was entirely restricted to the hippocampus. There was no damage to the surrounding cortical areas (the limbic cortex or the medial temporal lobe). In all the cases the damage occurred very early in life and the subjects attended school where performance was relatively good. They could not remember anything of their daily experiences, but, as evident from the fact they did reasonably at school, they were able to gain a vast amount of abstract knowledge. This study by Vargha-Khadem et al. suggests that damage to the hippocampus only affects episodic memory, but leaves semantic memory unscathed.³

So far I have cited one case that gives support to Shanks & St. John’s assertion that amnesiacs suffer from a failure of episodic, but not semantic, memory. But the case described above is not exhaustive of the variations that can be found between patients who suffer from amnesia. There is good reason to think that a disassociation exists between episodic and semantic memory, so it is no surprise that some subjects can be impaired in their ability to gain episodic knowledge, but not semantic knowledge. To show that amnesiacs do not meet (3) it would have to be the case that amnesiacs, in the majority of cases, only have impairment to episodic memory. As the following study shows, this is hardly the case.

In support of Vargha-Khadem et al.’s study, Verfaellie, Koseff & Alexander (2000) studied two amnesiacs that had contrasting deficits. Patient SS suffered from bilateral damage to the medial temporal lobe. This included damage to the anterior hippocampus, amygdala, entorhinal and perirhinal cortices leading to the septal regions and insular cortex (Verfaellie, Koseff & Alexander, 2000: 486). SS, who before suffering the damage to his medial temporal lobe was of above average intelligence, showed poor performance in the experiments, which tested for evidence of acquiring new knowledge. In contrast to SS, patient PS suffered only from a reduction in volume to the hippocampus, and performed far better than SS on tests for the acquisition of new knowledge. Both SS and PS exhibited severe failure of episodic memory.

By way of illustration, here are some of the results of Verfaellie, Koseff & Alexander. The first set of tasks tested to see whether or not SS and PS could remember words that had been introduced into use during different periods of time after the onset of amnesia. The recall task required the subjects to provide definitions for the words, while the recognition task had them select definitions from a list of plausible candidates. For the recall task SS’s average score was 16.3% compared to a mean score of 83.5% (SD=15.9) for controls. On the recall task PS did not fair much better than SS, with an average score of 23.9% while the mean for controls was 70.3% (SD=18.1). On the recognition task, again, SS did poorly with an average score of 30.4% (14/46, with chance cut-off at 17/46). However, on the recognition test PS fared much better with a score of 68.8% (22/32, with chance cut-off at 13/32), but was still impaired relative to the 90.8% (SD=4.2) of controls.

In an experiment that tested for knowledge of faces of people who had become famous after the onset of amnesia PS again fared better than SS. On the fame judgment task SS performed just at the chance cut-off 61.7% (37/60) compared to a score of 82% (SD=6.4) for controls. On the face identification task SS scored 10.2% to the 61.7% (SD=12.5) of controls. PS scored above chance on the fame judgment task with an average of 75% (27/36, with chance cut-off 24/36) compared to the 87.2% (SD=8) of controls. Like SS, the result of some sort of compensation, not seen in adults who suffer damage to the hippocampus later in life. Though in another study by Baddeley, Vargha-Khadem & Mishkin (2001) that involved one of the three subjects, results were consistent with those of Vargha-Khadem et al. (1997).

---

2 For the whole of this paper I am only discussing subjects with anterograde amnesia.

3 Tulving & Markowitsch (1998) discuss the implications of Vargha-Khadem et al.’s study. But see Zola, Squire, Teng, Stenfanci, Buffalo & Clark (2000) who suggest that the presence of semantic memory, in the absence of episodic memory, might be...
PS did poorly on the face identification task: 15.3% compared to controls: 68.1 (SD=12.2).4

This summary of the work of Verfaellie, Koseff & Alexander shows that amnesiac subjects can vary in how their memory is affected by brain damage (the variation seeming to depend on whether it is the hippocampus or surrounding cortical areas that have been damaged). Obviously, it is not paradigmatic of subjects that they suffer from a failure of episodic, but not semantic, memory. This can be seen by the fact that SS was impaired regardless of whether the task required recognition or recall, which is in contrast to PS for whom only recall was impaired.

Indeed, amnesic subjects who suffer from deficits to both episodic and semantic memory have been studied for quite some time. In a study of subject HM Gabrieli, Cohen, & Corkin (1988) found that he showed no knowledge of words that had been introduced since the onset of his amnesia. HM suffers from severe damage to the hippocampus and surrounding cortical areas: making his case of amnesia similar to that of SS.

Thus, even if one is unconvinced by the results of Verfaellie, Koseff & Alexander one can easily point to other results to make the same point I am making here: namely, it seems that amnesic subjects can experience a deficit in both episodic and semantic memory. Most likely this would be from damage to the hippocampus and other cortical areas. Thus, in SS we have a case of an individual who fits the criteria for scenario (3): someone who would fail to recall a study episode, and fail to exhibit evidence of semantic knowledge of that episode. Therefore, Shanks & St. John have not yet proven that amnesics fail to exhibit implicit learning.

Partial vs. Complete Disassociation

In response to commentary that evidence from work on amnesia is pertinent to the debate over the implicit/explicit learning distinction Shanks & St. John give a different argument against the relevance of amnesia. Several studies have found that amnesic subjects can experience a deficit in both episodic and semantic memory. Most likely this is from damage to the hippocampus and other cortical areas. Thus, in SS we have a case of an individual who fits the criteria for scenario (3): someone who would fail to recall a study episode, and fail to exhibit evidence of semantic knowledge of that episode. Therefore, Shanks & St. John have not yet proven that amnesics fail to exhibit implicit learning.

According to Shanks & St. John, it is the second of the above arguments that is relevant to implicit learning experiments. If it is the case that amnesics are impaired on the implicit test as well relative to controls, then the argument does not succeed. This is because it opens the possibility that the superior performance of controls on the implicit test is due to the extra explicit knowledge the controls have access to. As Shanks & St. John go on to argue, in most cases, implicit knowledge is either certainly impaired, or it is unclear whether or not it is impaired (1994: 436).

It is certainly to be expected that both PS and SS would be impaired on tests of implicit learning. And I would agree that one would be hard pressed to find a study where subjects performed at the same level as controls. However, there are two responses that can be made to Shanks & St. John. First, it is far too stringent to expect subjects to perform at the same level as controls on an implicit test, which Shanks & St. John expect (1994: 436). In fact, even if a clear dissociation of implicit and explicit learning was shown to exist, one should expect some impairment to the learning system that was not directly damaged. This is because any two systems that would operate as closely as two learning systems would inevitably be affected to some degree if one of the systems were severely damaged.

Consider olfaction and gestation. As a child, you might remember that if you held your nose when forced to eat something you disliked (I hated spinach) it wouldn’t taste as bad. Let us assume that when I was a child I could hold my nose very tightly, causing almost complete obstruction of olfaction. I would still taste the spinach pretty well, but it wouldn’t taste nearly as bad as it would if I had not held my nose. Plugging my nose is analogous to the sort of implicit and explicit tests that Shanks & St. John are criticizing: one system is severely impaired causing a lesser impairment to another system. Yet it does not seem warranted to conclude that olfaction and gestation can be explained by the same system, even though they are closely related. Likewise, it seems unwarranted to conclude that because amnesiacs are impaired on implicit tests, implicit and explicit learning can be explained by one system.

---

4 See Verfaellie, Koseff & Alexander (2000: 488-491) for methods, results, and discussion.
Second, following some commentators (Nagata 1994; Reber & Winter 1994), we can respond to Shanks & St. John in the following way: if a greater amount of explicit knowledge is available to controls compared to amnesiacs, and the same amount of implicit knowledge is available to both controls and amnesiacs, then a disassociation has been shown. The reasoning is that if both kinds of tests were of the same learning system, then we would expect amnesiacs to score lower on implicit tests than they actually do. The large discrepancy on the explicit test cannot be reconciled with the small discrepancy on the implicit test if we try to explain the results using a single system.

Shanks & St. John’s response is that the scores from explicit tests are not comparable to scores on implicit tests (1994: n.1). Just because percentages are used to represent scores on both tests, it does not follow that we can compare these percentages across tests. I think that this response misses the point.

As an example, let us again use Milner’s mirror-drawing test. Suppose that I was brought in as a control whose results would be compared to those of HM. In this case, let the explicit test just be a series of questions that are meant to test episodic memory and semantic memory for the learning phase: I score a 90% on this test. Suppose that the implicit test is just to see if I retain my ability to do the mirror-drawing task: I am given a grade of A. On the recall test HM does dismally, with a score of 52%. On the implicit test HM receives a grade of B+, which is comparable to my A. Shanks & St. John charge that it would be wrong to compare HM’s score of 52% to his grade of B+, and I agree. For one thing, the level of chance is no doubt different, not to mention other possible differences. But this is not what the commentators suggest: rather, we compare HM’s score of 52% to my 90%, and we compare HM’s B+ to my A.

In fact, the grading on the implicit test could be done using any form of measurement for differentiating between performances: it is unimportant. What is important is the difference that is found on a single test between subjects and controls. In the mock experiment HM does far worse than I on the explicit test, however, HM performs almost as well as I do on the implicit test. If we compare HM’s results to mine, then his explicit knowledge is far more impaired than his implicit knowledge. This is all that is needed for a disassociation between explicit and implicit learning systems to be shown to exist.

**Connectionist Models of Amnesia**

More recently, Kinder & Shanks (2001, 2003) have argued that, since they can create connectionist models that mimic results from work on amnesia, a single system can explain the apparent disassociation between explicit and implicit learning. In part one of this section I summarize their work, which I follow with my response in part two.

**Recognition, Classification, and Priming**

Several studies (Knowlton, Ramus & Squire, 1992; Knowlton & Squire, 1994, 1996) have found that in AGL experiments amnesiacs perform as well as controls on classification tasks. Thus, such studies support evidence from other research that posit distinct learning systems. Kinder & Shanks (2001) present a Simple Recurrent Network (SRN) model that achieves the same test results as those achieved by amnesiacs in AGL experiments.

In their model, Kinder & Shanks assumed that recognition and classification are functionally the same: that one system is responsible for both recognition and classification (2001: 649). Following this assumption, Kinder & Shanks manipulated the model by changing the learning rate to account for the general differences in memory ability between amnesiacs and controls. Learning rate affects the whole of a SRN: if I set the learning rate at x, then the learning rate of each individual node in the network is x. Thus, one parameter is manipulated and changes the system at a global scale. Kinder & Shanks conclude that a single learning system can account for the evidence found from research on amnesia. Since their SRN model performed in the same way as amnesiacs do on AGL experiments, and the SRN model involves one learning system, there is no reason to posit distinct learning systems.

Kinder & Shanks (2003) again did a similar study, which sought to mimic the evidence found for a disassociation between recognition and priming suggested by research on amnesia. Kinder & Shanks followed the experiments of Keane, Gabrieli, Mapstone, Johnson & Corkin (1995), which found amnesiacs to be impaired on recognition task, but not on priming tasks. Again, Kinder & Shanks assumed that amnesia is characterized by a general learning impairment and manipulated the learning rate in their model in both tasks (2003: 738-739). Again, Kinder & Shanks found that their results matched those of the study they targeted.

Given such findings, Kinder & Shanks held that their studies undermine the support that research on amnesia has given the implicit/explicit learning distinction. Most have thought that given the poor performance of amnesiacs on recognition tasks, compared to their performance on classification and priming tasks, studies on amnesia provide support for the distinction. But, since SRN models can produce the same results as amnesic and control subjects learning can be adequately explained using a single learning system.

**What SRN Models Do Not Prove About Learning**

Nothing about Kinder & Shanks’ models allow them to infer that they have undermined the support that research on amnesia gives the explicit/implicit learning distinction. We can assume it is true that their models’ results match those of Knowlton, Ramus & Squire and others, but this is only half the battle. Connectionist models can achieve any pattern of results, if the model is large enough, or parameters are manipulated in the right way. But there are two conditions that a connectionist model must meet if we want to conclude that the model accurately mimics some collection of results, let alone can be said to be the sort of system found in humans:

---

5 This is note one in the authors’ response to commentary.
1. The model’s performance is identical to the performance of the subjects it is designed to mimic in its performance.

2. The model achieves the same results as the subjects it is designed to mimic by being manipulated in a manner analogous to that which caused the subjects’ performance.

The SRN models of Kinder & Shanks seem to meet the first condition, or we can assume that they do, since it is the second condition that their studies fail to meet. What we are interested in is whether or not a disassociation can be found between implicit and explicit forms of learning. Thus, if we are to investigate certain deficits that are found in some individuals, then we are interested in the deficits themselves and in what appears to have caused those deficits. In the studies mentioned earlier, depending on whether damage is to the hippocampus or surrounding cortical regions, one seems to find different deficits.

If I want to explain a deficit that is caused by lesions it is no help to make a model that achieves the same deficit if the model is manipulated at a global scale. To explain a cognitive deficit with a connectionist model it is not enough to just have it mimic the performance of subjects with the deficit: you also have to mimic the cause of the deficit.

It is widely known that amnesia usually involves some sort of lesion, and not global neural change in some area of the medial temporal lobe. One could make a SRN model that mimics the performance of SS and PS, but if the model relies on manipulating the learning rate, then it has failed to explain the deficits of SS and PS. Kinder & Shanks (2001) briefly address the idea that a more “straightforward” way of trying to induce amnesia in the model would be to directly lesion it (2001: 663). Kinder & Shanks claim that, by removing part of the hidden layer and copy units of their model, they produced results that were similar those found with amnesiac subjects, but Kinder & Shanks do not report these results.

However, even assuming that such lesions did cause the SRN model to perform exactly like an amnesiac subject Kinder & Shanks still would not have shown that their single system model explained the deficits of amnesiacs. It is not enough to lesion the model, since the lesions should be similar to those found in amnesiacs. It is of course too much to expect Kinder & Shanks’ models to be organized like the hippocampus and surrounding regions. But in order for Kinder & Shanks to draw their strong conclusion, it would have to be the case that they manipulated their models in a way that was at least similar to the sorts of lesions found in the amnesiacs whom Knowlton, Ramus & Squire and others studied.

In the studies that Kinder & Shanks targeted amnesiac subjects did not suffer from global damage to learning centers. Knowlton, Ramus & Squire (1992) tested 13 amnesiac subjects: 6 of the subjects suffered from diencephalic amnesia (from Korsakoff’s syndrome, thalamic infarction, or penetrating brain injury), and the other 7 from confirmed (or expected) damage to the hippocampus. In all cases, lesion to the diencephalon or hippocampus was responsible for the amnesia (Knowlton, Ramus & Squire, 1992: 173). In fact, since the results of Knowlton, Ramus & Squire were obtained using amnesiacs that suffered from lesions to different brain areas, the results would only be accurately mimicked if the manipulations Kinder & Shanks made on their models were changed to account for lesion groupings. One has to do more than manipulate a single parameter to accurately simulate their results.

Keane et al. (1995) studied two subjects, LH and HM. LH suffered from bilateral occipital lobe lesions, while HM suffers from bilateral medial temporal lobe lesions (to the hippocampus and surrounding regions). As is usual, HM was not impaired in visual priming tasks, but LH was impaired on the visual priming tasks.7 Obviously the subjects studied by Keane et al. did not suffer from global deficits in learning in the brain, since lesions caused their deficits.

Given the discussion above, I see no prevailing reason why research on amnesia cannot be used as evidence of a disassociation between explicit and implicit learning systems.

Acknowledgments

Thank you to Jason Leboe, Tim Schroeder, and three anonymous referees for helpful discussion of and/or comments on earlier drafts of this paper.

References


Knowlton, B., & Squire, L. (1996). Artificial grammar learning depends on implicit acquisition of both abstract

---

6 Knowlton and Squire (1994, 1996) studied the same subjects as Knowlton et al.

7 LH suffers from Alzheimer’s, but unlike most subjects with this condition, he has a deficit in visual priming and not in conceptual priming. See Keane et al. (1995).

2030


