CHAPTER 4

*Implicit versus Explicit Learning*
The goal of this chapter is to arrive at a theory of implicit and explicit learning without introducing new theoretical entities. The basis for this theory will be the ACT-R architecture. The ACT-R theory, of course, also uses multiple theoretical entities. As we will see, none of these correspond directly to the notions of implicit and explicit learning, but together they can provide an explanation. This chapter will start with a general discussion about implicit and explicit learning. One experiment that is often quoted in the context of implicit learning is a dissociation experiment by Tulving, Schacter and Stark (1982). An ACT-R model is presented that can be used to explain their results. The model also serves as a basis for a more general discussion on how implicit learning and explicit learning can be understood in terms of ACT-R. The remainder of the chapter is used to discuss a particular example of explicit learning: rehearsal. Rehearsal is often studied using the free-recall task. By examining free-recall in several different situations, we may conclude that the primacy effect is mainly an effect of explicit learning, while the recency effect can be explained by implicit learning.

4.1 Introduction

In chapter 1 I have discussed Alan Newell’s criticism of psychological research, in which he mocked the simplistic conceptualization of the complexity of human cognition in terms of binary oppositions. Since 1973 a new opposition has become popular in cognitive psychology: the distinction between implicit and explicit learning or implicit and explicit memory. Although the term implicit memory was already proposed by Reber in 1967, the topic became popular by the end of the eighties. Before implicit learning research became popular, most memory research paradigms were based on either recognition or recall. Both in recognition and recall, participants first have to study some materials, and are tested later on. These types of experiments offer many insights into the nature of human memory, but tend to bias theories of memory. For example, in the famous dual-store memory theory by Atkinson and Shiffrin (1968), a major role for storing information in long-term memory is attributed to rehearsal, the mental process of sub-vocally repeating information. The dual-store theory was able to explain many of the recognition and recall experiments. A very powerful but false prediction was however neglected: the fact that no rehearsal implies no storage in long-term memory. As we will see shortly, learning may even take place without awareness. The dual-store theory overestimated the importance of rehearsal as a memory process, because it used recognition and recall as a basis. In both types of experiments, participants were told explicitly they had to memorize certain items.

Reber’s 1967 experiments departed from this experimental paradigm, and investigated what people learn without being aware of what they have to learn. The experiment he introduced, and which has been replicated many times in many
variations, is artificial grammar learning. In this experiment participants first study a list of strings that has been generated by a finite-state automaton based on an artificial grammar. After this study phase, participants were told the strings they had studied were words generated by a grammar. In the following test phase, they were presented with new strings generated using the same grammar, mixed with random strings and strings with subtle errors in them. Participants had to figure out which new strings were generated by the grammar, and which were not. It turned out that participants are surprisingly good at this task, and classify the new strings not perfectly, but well above chance level. Since none of the strings that were originally memorized were presented in the test phase, and participants were not aware of the fact that there was any systematicity in the learned strings, they somehow must have learned more than just the literal strings. Reber coined the term implicit learning to describe this additional, unintentional aspect of learning. Additional studies show that although participants perform well on this task, they can not explicitly state the rules of the grammar.

The idea that participants must learn to predict the behavior of a final-state automaton has been used in several other research paradigms. An example of one of these paradigms is dynamic system control, in which participants have to learn to control a complex system. An example is an experiment by Berry and Broadbent (1984), which involves a scenario in which participants have to learn to control a sugar factory. The Sugar Factory computer simulation they used is a dynamic system in which participants have to control sugar production by setting the number of workers. Since the relationship between input and output is highly non-linear, it is almost impossible for participants to discover the rule that governs the system. Nevertheless participants learn adequate control quite quickly, although they are not able to state the underlying rules of the system. A model of this experiment will be discussed in chapter 6.

Another type of research that deviates from traditional memory research is the dissociation paradigm. An example of this type of research is an experiment reported by Tulving, Schacter, and Stark (1982). In this experiment participants first had to study a list of 96 words. They were subsequently tested using two different tests, an implicit and an explicit test. The first, explicit, test was a simple recognition test, in which the participant was asked whether or not a certain word was in the study list or not. The second, implicit, test was a word-completion task. In this case participants were presented with a word fragment which they had to complete, for example A_ _ A _ _ IN (answer: ASSASSIN). Some of the fragments originated from the studied list, and others were from words not previously studied. Each participant had to do each test twice: an hour after the study phase and a week after the study phase. Figure 4.1 shows the results. One hour after studying the words, participants recognize 58% of the items correctly (this percentage is corrected for guessing). After a week, performance has dropped considerably to 24%. The implicit word-completion task, however, shows a totally different picture. Studying words
improves performance on this test: after one hour word-completion was accurate for studied words in 49% of the cases, while new words were only completed successfully in 30% of the cases. This advantage does not degrade with time, since after a week performance on the word-completion task is still the same. The discrepancy between the two tasks is called a dissociation: while one type of information, the fact that a word has been studied in the context of the experiment, degrades over time, other, subtly different, information seems not to suffer from any decay in time at all.

In the example above the dissociation is caused by time: one type of performance did suffer due to the passage of time, while another did not. There are other types of dissociations, for example due to brain damage. A study by Warrington and Weiskrantz (1970) reveals that patients suffering from amnesia perform much worse compared to healthy people on explicit tests like recognition and recall. On implicit tests like word completion, their performance equals control participants.

What do experiments such as artificial-grammar learning and dissociation learning exactly prove? At least they show the inadequacy of the classical recognition/recall paradigms, and also show that the “no rehearsal no learning” prediction of the dual-store model does not hold. But, probably to Alan Newell’s horror, psychologists turned the new phenomena into a new binary opposition, and, even worse, posed two binary opposite theories (the systems and the processing theory) to explain the distinction. Implicit and explicit learning were proposed as two distinct types of learning, each having its own mechanisms and needing its own theoretical framework. Explicit learning was associated with all the old memory research, but implicit learning, the new kid on the block, promised to be a new unexplored domain of countless experiments.

What makes implicit learning different from explicit learning? The dissociation experiments show that implicit learning is somehow more robust than explicit learning, since neither brain damage nor the passage of time seems to affect it.
Implicit learning is more robust in other aspects as well. McGeorge, Crawford and Kelly (1997) have shown that explicit learning is dependent on age and intelligence, while implicit learning is not. Participants that score higher on an IQ-test also perform better on explicit memory tests, and performance of older participants on the explicit test is worse than the performance of younger participants. Implicit learning on the other hand is hardly affected, either by IQ or age.

Another aspect of implicit learning, even used by some researchers as the defining quality, is that consciousness or awareness does not seem to play a role in it. Implicit learning is therefore sometimes called unconscious learning, as evidenced by the fact that although the participants can not verbalize any knowledge about the task, their performance increases nevertheless. In Reber’s artificial grammar participants were not able to state any of the grammar rules, but could categorize the strings anyway. In the Tulving experiment, participants had forgotten that they had studied a particular word a week after, but managed to use them for word-completion anyway. The notion of consciousness is, however, not unproblematic, as pointed out by Shanks and St. John (1994). In the artificial grammar experiments participants were not able to express any of the rules of the grammar. But they were aware of the fact that certain combinations of letters were more likely in grammatical than in ungrammatical strings, something that could at least explain some of their increased performance. A “safer” version of the unconsciousness aspect of implicit learning is to define implicit learning as unintentional learning, learning that is not tied to goals. In artificial grammar learning and in the Tulving experiment, participants had to memorize words or strings for later recall or recognition, not with the intention to do word-completion or to figure out a grammar. In this sense implicit learning can be seen as a “by-product” of normal information processing, while in explicit learning information processing is aimed at learning, comprehending or memorizing something.

There are two opposing theories that attempt to explain the differences between implicit and explicit learning: the systems theory and the processing theory. According to the systems theory, put forward by Squire (Squire & Knowlton, 1995), there are two different memory systems, an implicit and an explicit memory system, represented in separate structures in the brain. The fact that amnesiacs perform worse than controls on explicit tasks but not on implicit tasks can simply be explained by the fact that their explicit memory is damaged but their implicit memory is intact. Explicit memory is conscious memory, implicit memory is unconscious. Information in explicit memory decays with time, while information in implicit memory stays put. This also corresponds well with the folk-psychology idea that all our experiences are stored in unconscious memory.

The processing theory of implicit learning by Roediger (1990) assumes that there is a distinction between two types of processes: data-driven processes and conceptually driven processes. Data-driven processes are triggered by external
stimuli and can be associated with tests of implicit memory. For example, in the word-completion task part of the pattern is given. This part of the data actively facilitates the retrieval of the whole pattern. In the recognition test on the other hand, a connection between a word and an episodic event must be verified, so has a more conceptual nature. Conceptually driven processes are initiated by the participant and lead to explicit learning. According to the processes theory, memory performance will be best if the processing required on the test is the same as the processing required in the learning phase.

The problem with both the systems and the processing theory is that a distinction found in empirical data is explained by proposing two different theoretical entities, either two systems or two types of processing. From a scientific point of view this is a weak explanation that furthermore offers no insights in what the difference is between implicit and explicit learning. The evidence for separate entities is not final either. There are many examples of dissociations in which explicit learning is impaired while implicit learning is intact. If each type of learning is associated with its own theoretical entity, however, a so-called crossed double dissociation has to be found. In a crossed double dissociation, two experimental variables have to be found that have opposite effects on the implicit and the explicit test. A dissociation like this has never been found (Cleeremans, Destrebecqz & Boyer, 1998). To quote Cleeremans (1997, page 215):

With the exception of Hayes and Broadbent (1988) that has failed to be replicated so far, such a [crossed double dissociation] has never been observed in implicit learning situations. [...] the fact that no crossed double dissociation has ever been satisfactorily obtained in implicit learning research has often been used by other authors (e.g. Shanks and St John, 1994) as an argument to deny the existence of implicit learning as an independent and autonomous process.

Evidence from studies with patients isn’t strong either: both patients of Huntington’s disease (Heindel, Butters & Salmon, 1988) and Parkinson’s disease (Saint-Cyr, Taylor & Lang, 1988) have severe difficulty in learning motor skills, while showing intact performance on recall and recognition. Motor skills are usually considered procedural skills. Since people do not have conscious access to their procedural skills the associated learning process can be considered implicit. The problems these particular patients have, however, seem to limit themselves to the motor domain, so a generalization to implicit learning in general is unwarranted.

4.2 A model of the dissociation experiment

Tulving’s dissociation experiment consists of three separate activities, each of which is modeled by a small set of production rules: studying the list of words, the recognition test and the word-completion test. First, the list of words has to be
studied. In the experiment, every 5 seconds a word is presented. Since participants were only told they were involved in a memory experiment, they had no direct clue on what they had to do exactly with the words. It is therefore a safe assumption that participants will just rehearse the word and the fact that they have seen the word in the current context. This is easily accomplished in ACT-R: a first production rule creates a declarative recognition chunk that points to the word to be studied and to the current context. The recognition chunk can be considered as an episodic memory trace. A second rule keeps retrieving the chunk that represents the word and the recognition chunk until the next word is presented. Due to ACT-R’s base-level learning, the activation of a chunk is increased each time it is retrieved. The base-level activation at a certain time \( t \) can be calculated using the following equation:

\[
B_j(t) = \log \sum_{j=1}^{n} (t - t_j)^d + B
\]

In this formula, \( n \) is the number of times a chunk has been retrieved from memory, and \( t_j \) represents the time of each retrieval. The longer ago a retrieval was, the less it contributes to the activation. \( B \) and \( d \) are constants. Figure 4.2 shows an example of the behavior of this function, in which the activation of a chunk is plotted that is accessed at time 1, 4 and 7.

When the rehearsal production rule retrieves the recognition chunk and the chunk that represents the word itself, activations of both chunks are increased considerably, because \( n \) is increased in the formula, and the new \( t_j \)'s are all still close to \( t \). There is, however, a difference between the activation of the recognition chunk.
and the word chunk. The recognition chunk has just been added to declarative memory, so has no previous history of activations. This means that the activation of the recognition chunk is based solely on the few rehearsals in the context of the experiment. The word chunk, however, was already present in declarative memory, and already has a history of past use. In the model, this is simulated by assuming that words have been accessed on average 150 times, spread evenly over the past ten years, producing a low, but stable activation value. Some fixed activation noise in the model assures that all words have slightly different activation values. The difference between recognition and word chunks means that activations will also develop differently in the time period after studying the words. As figure 4.3 shows, both the word chunks and the recognition chunks start at a high level of activation. The activation of recognition chunks, however, decays faster due to the fact that they have no previous history.

In the recognition test the question must be answered whether or not a particular word has been studied in the study phase. In terms of the model this means that given a particular word chunk and a particular context chunk, a recognition chunk must be retrieved that connects the two. This is handled by two production rules. The first rule tries to retrieve the recognition chunk and answers “yes” when it succeeds. The second rule, which may fire if the first rule fails, just answers “no”.

This model is not entirely faithful, since it does not model the event in which a word that has not been studied is mistaken for one that has been studied. This can be modeled in ACT-R using partial matching, but this has not been done in the current model (partial matching has briefly been introduced in chapter 2, but will used in the Sugar Factory model in chapter 6). Failure to recognize a word that has been studied is due to the fact that the activation of the recognition chunk has become too low, since ACT-R cannot retrieve chunks with activations below the retrieval threshold.
A model of the dissociation experiment

In a recognition test, the indices to retrieve the right chunk are clear enough: the word and the study event. This is not the case in the word-completion task, where only a part of the word is given and the rest has to be retrieved. In order to retrieve the word that fits the pattern A _ _ A _ _ IN, ideally a production rule is needed that matches the first, fourth, seventh and eighth letter, and tries to retrieve a word that fits. The problem with this solution is that a production rule is needed for any combination of letters, which would mean 256 production rules if we would restrict ourselves to just 8 letter words. A solution that only requires a few production rules is to retrieve a word using only one or two letters, and compare if the retrieved word matches the rest of the letters. If it does, a solution has been found, if it does not, the model gives up. Alternatively, the model might have a few tries before giving up, but that aspect has not been modeled. One of the matching rules is as follows:

IF

the goal is to complete a word fragment AND
the first letter of the fragment is I1 AND
the second letter of the fragment is I2 AND
there is a word w that has I1 as its first letter AND
has I2 as its second letter

THEN

mark w as a candidate solution in the goal

This rule tries to find a word that matches at least the first two letters of the pattern. This rule will not work for the A _ _ A _ _ IN, because the second letter is unknown, but it will work if the first two letters are given.

Although both recognition and word completion require some declarative retrieval, they differ with respect to the source of errors. In the recognition test, it may be the case that a recognition chunk is no longer retrievable due to low activation. In the word-completion test interference with other words is the major source of errors. Words that are primed in the learning phase of the experiment get an activation advantage over words that are not primed. This advantage may persist over longer periods of time, as is indicated in figure 4.3. This difference between the two tasks may well be the real explanation for the dissociation. Figure 4.4a demonstrates that the model indeed behaves in a way that is comparable to human data. The main parameter that was manipulated to achieve the fit is the base-level learning decay (parameter \(d\) in equation 4.1). The recommended value for this parameter is 0.5, but this turned out to be a poor choice to explain long-term learning, since in a week ACT-R had forgotten everything. Instead the value of 0.3 has been used. Other parameters that have been manipulated, such as the retrieval threshold and the activation noise, did have small effects on the actual values of data points, but did not change the main dissociation effect.
The interesting aspect of this model is the fact that although it exhibits a dissociation, it nevertheless has no separate theoretical constructs to explain this difference. Both types of information are represented in the same memory system by the same memory process. The dissociation can be explained by the characteristics of the tasks themselves, rather than by hypothesized constructs. What is the difference between recognition and word-completion? To get a broader view on this question, we first have to review the notion of activation. Activation in ACT-R is an estimate of the log odds that a certain chunk is needed in the current context. This estimate is used in ACT-R for two purposes:

- If there are two or more possible candidates for retrieval by the production rule that is currently matched, the candidate with the highest odds is chosen.
- If the odds of needing a certain chunk are too low, the potential gain of retrieving it is not worth the effort.

If we look at the study task the participants have to do, we have to compare it to the situation in which people normally read words. In normal situations, it is not useful to remember in which particular context a word has been read. It is, however, useful to keep track of how often a word is used or encountered, since high-frequency words are more important than low-frequency words. So, if someone read low-
frequency words in a normal setting, he would typically not remember the event of reading the word itself, and would probably only update the frequency information of that word. The Tulving experiment is not a normal situation, it is a memory experiment. In order to meet the, at that point, unknown criteria of the memory experiment, the participant intentionally influences the normal learning scheme by rehearsing information. Rehearsal in this context means: intentionally increasing the number of retrievals of certain chunks, thereby artificially increasing the odds-of-being-needed of the chunk. As a consequence, the recognition chunk that stores the information that the word has been studied can still be recovered one hour after the study phase. A unintended by-product of rehearsal is that the frequency information of the studied words is increased as well. Since low-frequency words are used, the extra retrievals due to rehearsal have a significant impact on this estimate. It is this frequency information that the word-completion production rules need in order to select candidates, and which can be used as an explanation why studied words are completed better than words that are not studied, even after a week.

In the previous discussion the important difference between normal situations and a memory experiment is intentionality. In the introduction I have already noted that intentionality might be a key notion in the discussion. In the next section I will explain how this idea can be worked out in terms of the ACT-R theory.

4.3 An ACT-R theory of implicit and explicit learning

In the introduction I mentioned intentionality might be a good starting point to understand the nature of the difference between implicit and explicit learning. An advantage of using intentionality is that it can easily be operationalized in terms of ACT-R. Intentionality in terms of ACT-R means: tied to a goal. In the case of learning words for later recognition, as in the Tulving experiment, the intention of the participant is to memorize the words. If we look at the learning mechanisms in ACT-R, none of them is principally tied to intentions. Although the base-level learning mechanism may be used in the context of a memorization goal, it is not its basic function. Its basic function is to keep track of the odds that chunks are needed, a function that is normally performed unintentionally and unconsciously. The same can be said about all learning mechanisms in ACT-R: they are at work all the time, and are basically not tied to intentions. In a sense all learning in ACT-R is implicit learning. This idea is consistent with other properties of implicit learning. Implicit learning does not change much by ageing, and individual differences are small. This is exactly what we want for basic mechanisms in an architecture for cognition, since it is a theory about what people have in common and not about what sets them apart. The fact that implicit learning is not easily impaired due to brain damage also favors the architectural mechanism view: the basic way the brain works shouldn’t change due to damage.
What is explicit learning? The position I would like to defend is that explicit learning is a form of implicit learning. But while implicit learning is a by-product of normal processing, explicit learning is the by-product of specific learning goals. Where normal processing would retrieve a chunk representing a word only once, an explicit learning goal may retrieve it a number of times, not because it is necessary for processing, but just to put the implicit learning mechanisms to work. Although we have no direct conscious access to the base-level learning mechanisms itself, we may have found out, due to experience that repeating a word helps remembering it. Instead of being another type of learning, explicit learning is just a set of strategies to make the best possible use of the implicit mechanisms. Explicit learning is therefore not a part of the architecture of cognition, but is rather produced by knowledge that is represented in the memory systems of that architecture. This idea also corresponds well with properties of explicit learning: since the knowledge corresponding to it has to be learned itself, one can expect large individual differences due to intelligence and development. Similar observations can be made with respect to brain damage. If implicit learning is a fundamental property of the brain, it will not be easy to damage it. Explicit learning, on the other hand, consists of knowledge. Brain damage may cause this knowledge to be lost, or disrupt successful usage of this knowledge.

In the case of the Tulving experiment, the recognition task is an explicit task only because participants suspect either recognition or recall if they are told they are involved in a memory experiment. If one explained the word-completion task to participants at the start of the experiment, and told them they were supposed to do this task after the study phase, it would turn into an explicit task. The participant has several options: she can either stick to a rehearsal strategy, or attempt some more clever memory strategy, for example by explicitly memorizing characteristic fragments of words. The choice of strategy will have a large impact on performance. The original rehearsal strategy will of course still exhibit the assumed characteristics of implicit learning, while the fragment-memorization strategy, if it works at all, will probably suffer from the same fast decay that is supposed to characterize explicit learning. We might even be able to find a dissociation within the same task in healthy participants.

In Reber’s artificial grammar and Berry and Broadbent’s sugar factory, participants’ performance increases, although they are not capable of formulating any explicit rule-like knowledge about the task. In both cases, it is very hard to find the real rules: deriving grammars from examples is a very difficult task, and the non-linear character, the randomness and the limited means of control in the sugar factory make it almost impossible for participants to derive rules within the limited time of the experiment. As a consequence, explicit strategies that are usually successful in detecting regularities will fail. Nevertheless there is also implicit learning going on. For example in the sugar factory task, which I will discuss in detail in chapter 6, each time the participant sets the controls of the factory and perceives an outcome, a chunk recording this information is added to declarative memory. This is not done
intentionally, but rather because all popped goals are stored. It will turn out that this information alone can account for the improvement participants show on the task.

In the remainder of this chapter and in the next two chapters, I will explore the implicit/explicit distinction based on the idea that implicit learning is based on mechanisms of the architecture, and explicit learning is the application of learning strategies. In chapter 5, I will discuss explicit strategies that learn new production rules, and how an increase in the number of strategies can explain the difference between small children and adults on a classification task. In chapter 6, I will describe how the implicit/explicit learning debate can be related to another debate in the learning literature: whether new skills are learned by accumulating examples, or by deriving general rules. The remainder of this chapter is devoted to one of the issues stated in the previous chapter: a model of rehearsal. This model will be discussed in the context of the free-recall task, a classical paradigm to study rehearsal.

4.4 A model of rehearsal and free recall

The model discussed in this section is the first model I made in ACT-R. As a consequence, the model is based on an old version of ACT-R (2.0), which on the one hand included features that have since been removed, but on the other hand did not include all that is currently part of ACT-R. I further chose to implement verbal rehearsal using a separate phonological loop, based on Baddeley’s evidence for this kind of structure. If I were to model rehearsal again, I probably would be more hesitant to add extra structures to the architecture. Recently, the CMU group (Anderson, Bothell, Lebiere & Matessa, 1998) has also modeled free recall as part of a broader project on list learning. Their model did not use an explicit phonological loop. They, however, implemented a phonological-loop-style memory structure within declarative memory that did the same job.

As we have seen in the introduction, rehearsal has been studied extensively in the seventies in the context of the dual-store memory theory by Atkinson and Shiffrin (1968). One of the experimental tasks used for studying rehearsal is the free-recall task. In this task a list of words, typically containing fifteen to twenty items, is presented at a constant rate to a participant. After presentation, the participant has to recall as many words as possible from the list. Two effects emerge from the results, the primacy effect and the recency effect, respectively referring to the fact that the first and the last few items of the list are recalled better than the rest. The dual-store memory theory can explain both effects: the primacy effect is due to the fact that the first few items in the list are rehearsed more often because they initially don’t have to compete for space in short-term memory (STM), and the recency effect is due to the fact that the last few items are still in STM at the moment they have to be recalled.
This explanation is confirmed by Rundus (1971), who asked participants to rehearse aloud. The data show that there is a relation between the number of rehearsals and the chance of recall (figure 4.5), at least with respect to the primacy effect.

Since the popularity of the dual-store theory declined, partly because rehearsal turned out to be not the sole mechanism to store information in long-term memory (LTM), less research effort has been put into it. A theory that does involve rehearsal is Baddeley’s theory of working memory (Baddeley, 1986). In Baddeley’s theory, working memory has a central executive and two rehearsal subsystems: the phonological loop and the visuo-spatial sketch pad (figure 4.6). Both subsystems are used to temporarily store small amounts of phonological and spatial information. The phonological loop is a system that stores up to two seconds of phonological code in a serial fashion. The visuo-spatial sketch pad uses a quasi-visual representation of objects that can be used for spatial reasoning. The visuo-spatial sketch pad can be used to answer questions like: if the triangle is below the square, and the circle is to the right of the square, and the circle is above the cross, is the cross left or right from the triangle?
A model of rehearsal and free recall

The phonological loop is the relevant structure for retention in free recall, at least in the overt-rehearsal version by Rundus. Instead of being the process that transfers information from STM to LTM, rehearsal has become a process necessary to maintain items in STM. Whether or not information will also be stored in LTM is not specified by Baddeley's theory, because it is a theory of working memory only. Work by Craik and Lockhart indicates that the extent to which rehearsed information is stored in LTM depends on the amount of processing that needs to be done on individual items (Craik & Lockhart, 1972). This led to the distinction between maintenance rehearsal and elaborate rehearsal. Maintenance rehearsal is used just to retain information for a short time, for example a telephone number that needs to be dialled. During elaborate rehearsal on the other hand further processing is done on the rehearsed information.

Baddeley has gathered extensive empirical evidence for the phonological loop and the visuo-spatial sketch pad. The central executive, however, is a weak point in the theory. It is supposed to be able to contain two or three items, and to control what goes into both subsystems, but it is unclear what representation it uses, and why and when it puts something in either subsystem. The central executive is almost a reference to the rest of information processing, because it not only stores information, it also makes important decisions on what to memorize in what subsystem. Some of these decisions must be deliberately planned, involving knowledge stored in LTM. The problems with the central executive have an obvious reason: somehow the theory of working memory must be tied to the rest of information processing, and the central executive is responsible for this.

The ACT-R theory can be seen as a specification of central information processing that can serve as a means to create models of rehearsal using Baddeley's phonological loop. The role of the central executive is taken care of by the ACT-R architecture.

A model of free recall in ACT-R
To be able to model free recall in ACT-R, we first need some way to do rehearsal. In order to use Baddeley's phonological loop, some assumptions have to be made about the representation of the loop and the interaction with ACT-R. According to Baddeley, the phonological loop has a phonological representation. To be able to interact with the memory of ACT-R, we must assume it is possible to activate a phonological representation given a chunk-like symbolic representation in declarative memory and vice-versa. To simplify matters, we will assume the phonological loop has the following properties:

- The phonological loop is a linear storage buffer with a capacity of 2 seconds of phonologically coded words.
- References to declarative chunks representing pronounceable words can be added to the loop. New references are added to the end of the loop.
4: Implicit versus Explicit Learning

- If the capacity of the loop is exceeded, a random word is dropped.
- At any moment the contents of the loop can be rehearsed, which involves entering a subgoal to do this.
- In the rehearsal subgoal the words can just be rehearsed (maintenance rehearsal), or further reasoning can be done with them (elaborate rehearsal).

Implementing a separate structure for rehearsal is at odds with the idea that rehearsal is just a learned strategy. But what if the phonological loop is not primarily a structure of working memory, but rather a buffer to store perceived speech in, or speech that is about to be pronounced? In that case, rehearsal would be a clever strategy of reusing a structure whose original purpose is different.

Once rehearsal is taken care of, a model of free recall is straightforward. During the study phase of the experiment words are read and added to the phonological loop one at a time. In the time between presentations the phonological loop is rehearsed. At the time of recall, words are recalled in order of activation until there are no words left above the retrieval threshold. No attempt is being made to first “empty” the phonological loop at the time of recall, only the last item of the list is retained.

The explanation this model offers for the two prominent effects in free recall, the primacy and the recency effects can now be made clear. The primacy effect can be explained in the same manner as Rundus’ explanation: the first few words are rehearsed more often, on average, so are retrieved more often. The recency effect can be explained by the fact that the retrievals are relatively recent, so their impact on the activation is larger.

A positive recency effect can be considered as an implicit learning effect, since its presence is not influenced by strategy. This finding concurs with developmental data. Hagen and Kail (1973) compared free-recall behavior of 7 and 11 year-old children. Although both groups show a recency effect in recall, in the group of younger children the primacy effect is absent. Cuvo (1975) found that this difference can be attributed to strategy: younger children tend to just repeat the last item presented, while older children adhere to the adult pattern of rehearsal. These studies demonstrate that implicit learning, as witnessed in the recency effect, is not affected by age, while explicit learning is, as witnessed in the primacy effect.

Simulation 1
The goal of the first simulation was to reproduce the results of Rundus’ experiment. Rundus used 25 participants, to whom 11 lists of 20 words were presented on cards with a 5 second interval. Participants were instructed to rehearse aloud.
In the experiment the mean number of words correctly recalled was 11.12 and the mean number of rehearsals 88.3. The simulation recalls 11.15 words correctly on average, using 116.0 rehearsals. The serial position curve and the mean number of rehearsals for each item in the list are shown in figure 4.7. The fit between the data and the model is reasonably good for the probabilities of recall ($R^2=0.82$), and not too good for the number of rehearsals ($R^2=0.57$). As can be seen in the figure, the model overestimates the number of rehearsals, although the curve has the right shape.

**Simulation 2**

In the standard experiment, participants have to rehearse aloud, but are free in choosing which words to rehearse. Participants can be constrained in this aspect, for example if they may only rehearse the word that has just been presented. Figure 4.8
shows the data (from Fischler, Rundus & Atkinson, 1970) and the results of the model ($R^2=0.65$). The interesting aspect is that the primacy effect largely disappears, but the recency effect remains. This finding is consistent with Hagen and Kail (1973) (no primacy effect in young children) and Cuvo (1975) (young children only rehearse the last word presented) studies.

**Simulation 3**

To see whether the model holds its ground in other variants of the task, a data set collected by Murdock (1962) is a good basis for comparison, since he used different list lengths (from 10 to 40 words) and different rates of presentation (1 or 2 seconds per word). Murdock did not require overt rehearsal, so only the probabilities of recall can be compared. Figure 4.9 shows the data and the results of the model. The main deviation between model and data is that the model overestimates the primacy effect. The overall explained variance is nevertheless quite high ($R^2=0.91$).

**Simulation 4**

In the standard free-recall experiment, recall starts immediately after the presentation of the words. If there is a delay between recall and presentation in which further rehearsal is prevented, the recency effect disappears. An experiment by Postman and Phillips (1965) demonstrates this effect: 18 participants were given lists of 20 words, 6 lists for which recall immediately followed the presentation, and 6 lists where participants had to count backwards for 15 seconds before recall. Words were presented at a rate of one word per second, and rehearsal was covert. The mean number of words recalled correctly was 6.20 if there was no delay after presentation, and 5.05 if there was a 15 second distraction. The serial position curves for both conditions are depicted in figure 4.10, together with the simulation data. The simulation recalls 8.6 words correct on average in the condition without...
delay, and 4.6 words in the 15 sec delay condition. The most interesting aspect, however, is that the recency effect has largely disappeared. This is normally explained by the fact that participants cannot use the contents of their rehearsal buffer in their answers, but the model shows that an explanation based on decay of activation is sufficient. It also predicts that due to the fact that the last few items are rehearsed fewer times than items in the middle of the experiment, the recency effect will eventually turn into a negative recency effect, as we will see in simulation 5.

The primacy effect is much less affected by the delay, since it is caused by the fact that items have been rehearsed more often. The explained variance is only average: the overall $R^2$ has a value of 0.58.

**Simulation 5**

Craik (1970) discovered that the disappearance of the recency effect after a delay can even turn into a negative recency: in some situations recall for items at the end of the list is worse than for items in the middle part. In a free-recall experiment 20 participants were presented with 40 lists of 15 words at a rate of 2 seconds per
word. After each 10 lists, participants were asked to recall as many words as possible from the previous 10 lists, giving a final-recall score. The results of this experiment are shown in Figure 4.11a. To obtain a smooth curve Craik averaged each data-point with its successor and predecessor, except for the first and the last.

The free-recall model also produces negative recency, as can be seen in figure 4.11b. The same averaging technique as Craik used is used on the data. In the simulation the model has to produce as many items as possible after presentation, after which a 60 second break follows and another, final, recall session. Although the results of the model cannot directly be compared to Craik’s data, since the experimental setup is different, a negative recency effect that is similar to Craik can be seen in the model.
Discussion

The results of the simulations show that the classical effects of primacy and recency in free recall can be reproduced using a theory of rehearsal based on the ACT-R architecture and Baddeley’s phonological loop. The primacy effect can be explained by the fact that items early in the list are rehearsed more often on average than other items in the list, the same explanation that was used in the dual-store theory of memory. The recency effect can be explained by the base-level activation mechanism of ACT-R: the last few items of the list have a higher activation because they have been accessed more recently.

Simulations 2, 4 and 5 show that both the primacy and the recency effect can be manipulated by changing aspects of the task. It is interesting to examine the nature
of these manipulations. In simulation 2, participants were instructed to use a certain type of rehearsal strategy, which resulted in the disappearance of the primacy effect. The learning strategy thus determines the presence or absence of the primacy effect, and can be considered as an effect of explicit learning. In simulations 4 and 5, the circumstances of the experiment were changed. Instead of changing the strategy, a time delay was used, resulting in an effect on the recency effect.

The various models presented in this section also illustrate the inadequacy of the $R^2$ measure to express the quality of fit between the data and the model. Although the fit with the original Rundus data is clearly the best, the model of the Murdock experiment achieves the best fit, although it overestimates the primacy effect.

The parameters in the models discussed above were set to their recommended default settings, except for the activation noise and activation threshold, which were estimated to optimize the fit to the Rundus model. The same settings were used for all the other simulations. The base-level learning decay parameter used was the recommended value of 0.5. In the Tulving model this parameter had to be set to 0.3, meaning there is an issue to be resolved here. We will return to this issue in chapter 6.