Forms of memory: Investigating the computational basis of semantic-episodic memory interactions
Neville, D.A.

Citation for published version (APA):
Chapter 5
The goal of the present chapter is to gain some insights on the mechanisms that might explain the behavioral findings reported in Chapters 2, 3 and 4 of this thesis. In the following sections, I will first discuss the empirical findings in light of current global models of memory and then discuss some theoretical considerations on the basis of how each model explains the data. The unifying thread for this theoretical discussion will be the identification of shared computational principles and their relevance for the functional interactions of semantic and episodic memory.

5.1 Global models of memory

The study of how different memory systems combine to produce behavior is often confronted with the problem of comparing alternative accounts (or theories) against the empirical evidence. From a modeling perspective, the comparison of different accounts of the data requires the description of how information is computed to generate behavior; in other words a formal model of the steps needed to transform available input (stimuli) in observed output (data). Formal modeling constitutes a unique tool for research on memory since it offers an appropriate analytical framework where different theories of memory functioning can be expressed and compared against empirical findings.

On first approximation, models of memory can be broadly classified in three major categories (Raaijmakers & Shiffrin, 1992) depending on the type of predictions they entail: qualitative, quantitative and semi-quantitative models. Qualitative models are cast at a very general level of description and focus on explaining whether performance differs across sets of conditions, tasks or manipulations. Quantitative models instead are cast at a more fine-grained level of description and focus on the precise prediction of experimentally observed quantities (e.g. RT). Semi-quantitative models fall in between these two classes and focus on predicting the right order of magnitude for effects of interest with, for example, a large effect in one task (i.e. priming of 1 sec) and a significantly smaller effect in another task (i.e. priming of 10 ms). It is important to note that the scale or metric by which an effect is deemed ‘small’ or ‘large’ is completely dependent on the model or task under examination, therefore what is informative for the semi-quantitative approach is whether a given model is capable of predicting effects in the right direction (i.e. positive or negative) and with the correct
magnitude (i.e. priming of 10 ms vs. priming of 10 sec).

Models of memory are very often developed initially to provide a consistent explanation of the findings from a single behavioral task or experiment. Some of these models however are later on generalized to accommodate findings from other tasks and experiments as well. Such generalized models are commonly referred to as global models of memory since they constitute a unified set of computational principles from which task-specific models can be derived, usually by introducing additional (task-dependent) assumptions. The core set of principles underlying a global model of memory represents thus the formal definition of a unified theory of memory processing.

Three theoretical frameworks have been extremely successful in the past years in explaining a wide range of findings related to episodic and semantic memory, namely ACT-R (Anderson, 1993), SAM-REM theory (Raaijmakers & Shiffrin, 1980, 1981; Gillund & Shiffrin, 1984; Malmberg & Shiffrin, 2005; Mensink & Raaijmakers, 1988; Shiffrin & Steyvers, 1997) and connectionist models (McClelland, 2000; McClelland & Rogers, 2003). Given the limited space of the present chapter I will focus on the specific models within these theoretical frameworks that are mostly relevant for the present thesis, namely the models eSAM (Sirotin, Kimball & Kahana, 2005), REM-I (Schooler et al., 2001) and REM-LD (Wagenmakers et al., 2004), RACE/A (Van Maanen et al., 2012c) and distributed models of semantic memory (Becker et al., 1997; Masson, 1995; Seidenberg & McClelland, 1989).

I will now briefly introduce each model in turn and highlight the core mechanisms that I will invoke in the second part of this chapter to discuss the findings of Chapters 2,3 and 4. In order to provide a principled discussion of the differences between global models of memory, each of them will be discussed in relation to three key questions: 1) how semantic and episodic associations are represented in memory, 2) how is context modeled in each memory system and 3) how different features enter the computations of memory processes.

5.1.1 The SAM and eSAM models

The SAM (Search of Associative Memory) model (Raaijmakers & Shiffrin, 1980, 1981) is one of the most influential global models of memory and it has been very successful in explaining a number of findings from both free recall (Raaijmakers, 1979)
and recognition (Gillund & Shiffrin, 1984) tasks such as the effects of list length and list strength, presentation time, serial position effects, the temporal aspects of recall, context effects on recall and recognition, interference and forgetting (Mensink & Raaijmakers, 1988), the part-list cuing effect (Raaijmakers & Phaf, 1999), spacing effects (Raaijmakers, 1993) and a number of dissociations between recall and recognition (Gillund & Shiffrin, 1984).

In the SAM theory the assumption is made that information is stored in memory in the form of "memory images" (or memory traces). Each image is the collection of three types of features: item, associative and contextual features. Item features reflect aspects that are specific to a single object represented in memory, for example the semantic meaning or the orthographical form of a word. Associative features instead describe how different items are mutually related to each other, for example in terms of associated semantic meaning (e.g. cow-milk). Lastly contextual features capture both the external and the internal environmental aspects related to the episode during which the memory was encoded (e.g. time of the day or emotional status). In this model the amount, type and quality of the information stored in an image is determined by the encoding processes that are deployed in short-term memory (e.g. elaborative rehearsal) and by the features being active at the time of encoding.

The next crucial assumption of SAM is that the retrieval of information from long-term memory is a cue-dependent reconstructive process. Cues can be words from the studied list, category cues, contextual cues, or any other type of information used by the subject in the attempt to retrieve a memory image from long-term memory. The outcome of the retrieval process, whether an image is retrieved or not, depends on the strength of the association between the retrieval cues and that particular image.

Regarding how semantic and episodic information can be represented in this framework, Sirotin, Kimball and Kahana (2005) recently proposed an extension of the SAM model, termed eSAM, which makes the additional assumption of structured semantic associations existing prior to experimental testing alongside episodic associations formed during the experiment. In eSAM both the pre-existing structure for semantic associations (part of the lexicon) and the structure for episodic associations are modeled as matrices, respectively termed, the semantic matrix (SM) and the episodic matrix (EM). Intuitively the SM matrix is used to capture the associations laid out in semantic memory between pairs of items that do not change over the course of
the experiment, in other words the pre-existing lexical-semantic knowledge of the subject. The SM is computed starting from a lexicon using some measure of semantic relatedness and various approaches can be used. For example in the simulation studies of Sirotin and colleagues a combination of latent semantic analysis (LSA) and word association space (WAS) was used to transform free-association norms (Nelson, McEvoy & Schreiber, 2004) into a multidimensional semantic space. The results of these computations were then stored in the semantic association matrix (SM) to model both the structure and the strength of the semantic associations found in the lexicon. The EM matrix (equivalent to the retrieval structure in SAM) instead is used to store both inter-item and item-to-context associations and models the formation of episodic associations during the experiment.

The next question of interest is how context is modeled. The classical view on context and on how it differs between semantic and episodic memory is that episodic memory is rich in contextual information whereas semantic memory tends to be insensitive to (episodic) context and abstract in nature (Tulving, 1972). As in previous versions of SAM, in eSAM episodic associations are incremented during study and recall as a result of information cumulating across multiple presentations. Context, as a cue, instead fluctuates over presentations due to sequential accumulation of different information across episodes. This entails that multiple presentations lead, on the one hand, to a positive increase in performance due to the strengthening of episodic associations. On the other hand, multiple presentations lead to a lack of increase (or possibly decrease) in performance due to the fact that context, as a cue, fluctuates constantly, therefore, never being strongly associated with any item in particular. In eSAM this change in the amount of context associated with a list (contextual drift) is modelled by assuming a probabilistic exponential decay of the item-to-context strength. The end result of the contextual drift mechanism is a simple auto-regressive component that was shown to be sufficient for most cases in the study of Sirotin et al. (2006), however see Mensink and Raaijmakers (1988) for a more complex form.

The third question of interest is how these assumptions come together in computational terms. In other words, how semantic, episodic and contextual features are processed during the sampling and recovery of information from long-term memory (LTM). In eSAM the probability of sampling an image from LTM is computed similarly to the general retrieval mechanism proposed by Raaijmakers and Shiffrin (1980); that is a
sampling rule that combines multiplicatively the strengths of semantic, episodic, and contextual associations. As in previous versions of SAM, use of a multiplicative retrieval rule implies that each type of association—semantic, episodic, and contextual—modulates the influence of the other types of associations on sampling probabilities. Therefore, in the scenario where two items have relatively high contextual strengths and high episodic strengths of association to a third item but one of them has a higher semantic strength of association to the third item than the other, eSAM predicts that the item with the higher semantic strength will be more likely to be sampled when the third item is used as a retrieval cue. Notably both of the items might have a higher probability of sampling than other items, because of their high contextual and episodic strengths, and this advantage may increase if episodic and contextual associations are given more weight during retrieval.

In summary the eSAM model introduces two important additions over the existing SAM model to capture semantic-episodic interactions: 1) explicit modeling of both semantic and episodic associations, 2) fluctuation of the context cue (see also previous work from Mensink and Raaijmakers, 1988). These two assumptions are then combined within the general probabilistic sampling mechanism of SAM theory to explain reciprocal influences of semantic and episodic associations on the retrieval from LTM.

5.1.2 The REM models: REM-I and REM-LD

A second global model of memory intimately related to SAM is the REM (Retrieving Effectively from Memory) model for recognition memory (Shiffrin & Steyvers, 1997). This model was initially developed to solve the problem posed by the mirror effect for global memory models. The mirror effect refers to the finding that many factors that increase the probability of a 'hit' (saying 'yes' to a target item) also decrease the probability of a 'false alarm' (saying 'yes' to a distractor item) (Glanzer & Adams, 1985; Glanzer, Adams, Iverson & Kim, 1993). Therefore, the order of the conditions for the probability of saying 'yes' to distractors is the mirror image of the order for these same conditions for the probability of saying 'yes' to target items. For example, even though low-frequency (LF) items are more likely to be correctly rejected than high-frequency (HF) items, LF target items are also more likely to be correctly
recognized than HF targets. Such mirror effects are difficult to explain for any model that bases the probability of saying 'yes' on a "strength"-like measure (as the strength of activation in SAM). Although mirror effects can be handled by assuming different response criteria for HF and LF items, such a solution is inelegant and somewhat problematic since it is difficult to find coherent explanations of how the criterion moves across conditions.

The REM model proposed by Shiffrin and Steyvers (1997) provides a more elegant and consistent explanation of the mirror effect by making an important assumption, namely that the memory system behaves as an optimal decision maker. Memory images are represented in this theory as vectors of feature values, e.g. \( <3,1,3,7,3,...,2,1> \). The numbers represent the frequency of a particular feature value. When an item is studied, an episodic image (of item and context features) is formed and stored in memory. Due to noise in the encoding processes, this episodic image will be error prone with some features being stored incorrectly and some others not being stored at all. In general the longer or more accurate the processing of an item during study, the higher the probability that a given feature will be stored correctly in memory. In a recognition test old ('studied') items are presented along with new ('novel') items and subjects are asked to indicate whether the presented item is an old one (from the list of the studied items) or a new one. REM theory assumes that the system compares the features extracted from the probe (i.e. presented item) to those stored in the episodic memory images. The crucial property of this comparative process is that both the matching and the mismatching features of each image are diagnostic for the decision, in other words, both provide information for choosing between the response alternatives (i.e. 'old' vs. 'new'). More technically, the system uses a rational, Bayesian, criterion for response selection by choosing the response that has the highest probability given the observed number of matching and mismatching features across all the memory images. An episodic image that is quite similar to the test item will produce several matching features that in turn will lead to a high probability that the test item is an old one. In mathematical form the rational decision criterion is expressed by the posterior odds ratio.

In the past years REM theory has been successfully applied to explain a number of findings for implicit memory paradigms such as perceptual identification (Schooler et al., 2001) and lexical decision (Wagenmakers et al., 2004). The distinction between
implicit and explicit forms of memory is usually drawn depending on whether the participants are aware of the fact that their memory is being tested. An example with the priming effect will further clarify this distinction. Repetition priming usually refers to the finding that recent study of a word enhances (or at least alters) performance on a subsequent generic memory test. In the explicit version, participants are informed before the final test (i.e., a recognition test) that the items to be presented are from a previously studied list. In the implicit version instead, participants are unaware that the some of the items presented during the test phase are from the previously studied list.

In order to explain repetition priming effects, REM relies on the assumption that during study there is not only the encoding of a new episodic image but also additional storage of contextual information in the (pre-existing) lexical/semantic image corresponding to the studied word. The additional assumption is made that the information being added to the memory image is restricted to information not already present in the lexical/semantic image. From these assumptions it follows that information already stored in the lexical/semantic image such as its core meaning is unaffected by recent study events whereas information which is unique to the current study episode, such as perceptual (e.g. font) and context information, is constantly added to the lexical/semantic image. Hence, the lexical/semantic trace is not static but is dynamic and sensitive to recent presentations.

Repetition priming effects are in this view a natural consequence of the storage assumptions, as long as the test probe utilizes any of the information (features) that has been added to the lexical/semantic image. If current context is used as part of the test probe, which may be inevitable even in tasks that do not require context cuing, then the match of this information to that stored in the lexical/semantic image of the studied word will increase the likelihood ratio for that image, and produce priming.

As discussed in Schooler et al. (2001), REM theory can predict priming effects in perceptual identification if one assumes that prior study leads to the storage of a small amount of new contextual information (i.e., prior study does change "some property of the representation"). The general idea is simple. The extra matching of context features for the studied item increase the likelihood ratio for choosing the target lexical/semantic image over all of the other alternative images stored in lexical/semantic memory. A similar explanation can also be formulated for semantic memory based tasks such as lexical decision (Wagenmakers et al., 2004). Similar to the case of perceptual
identification, the previous presentation of a given word produces an advantage of the target word item over the non-word distractor by means of an increase in matching (and mismatching) features. The rational decision process then evaluates matching and mismatching features for both word and non-word items. The additional context features stored during previous representations lead to an increase of the likelihood ratio for choosing the correct response (i.e. word) over the incorrect one (i.e. nonword) and therefore to a repetition priming effect for previously processed items.

In both REM-I and REM-LD models semantic-episodic associations are assumed to have separate representations, similar to the approach of eSAM. However, contrary to eSAM, context is assumed to be part of both episodic and semantic memory since context features are added to both episodic and (to a lesser extent) to semantic memory images. An interesting idea in this respect, which is not part of the original formulation of eSAM, is that the similarity of memory images might regulate the storage of context features with features being stored only when the similarity between the episodic trace and the semantic one is high enough. In both models semantic, episodic and contextual features enter the computations of memory retrieval via the rational decision process that evaluates matching and mismatching features between a given memory image and (all of) the alternative memory images. Positive (match) and negative (mismatch) information is (in virtue of the rational assumption) diagnostic for the decision process and therefore provides an explanation of repetition priming effects for both explicit and implicit memory tasks.

5.1.3 The ACT-R and RACE/A models

A third global model of memory that has been extremely successful in predicting declarative and procedural aspects of memory performance is ACT (Anderson, 1983b). Given the limited space of the present section, I will discuss only the latest version of ACT theory, ACT-R (Anderson et al., 2004). In this formulation of ACT theory, the cognitive system is assumed to consist of a declarative and of a procedural memory system. The procedural system is assumed to be a set of production rules of the form if stimulus $X$ is detected then implement response $Y$. The declarative system instead is represented as a large set of nodes interconnected by links. The nodes represent basic concepts or cognitive units (i.e. the concept of 'house') and the links represent semantic
or episodic relations. Whenever two items are studied together in a memory task, a link between the corresponding nodes may be formed. Information is stored in chunks and the retrieval from long-term memory is expressed as a function of the activation level of a chunk. The basic idea of ACT-R is that the cognitive system has developed in such a way as to provide an optimal or rational response to the informational demands of the environment. In such a model, the retrieval of a target item B from a cue item A is accomplished when the activation from the node representing A spreads to item B and activates the node representing B (or sends enough activation to B to pass an activation threshold). In the application of ACT-R to memory paradigms (see Anderson, Bothell, Lebiere & Matessa, 1998) the assumption is made that the activation of a chunk depends on both its base-level activation, a function of its previous use, and on the activation that it receives from the elements currently in the focus of attention. ACT-R can be seen as an extension of classical spreading activation theories of semantic memory (Collins & Loftus, 1975) however the crucial assumption is made that the spreading of activation over the network is not restricted by the spatial configuration of the network links (whether two nodes are connected or not). In other words the activation of a chunk is directly related via its association to the source elements.

An important similarity between ACT-R theory and SAM-REM theory is that the activation of a chunk can be interpreted as the odds that the information will be needed in the current context. Bayes’ rule provides a formal way to express this concept: the probability that a particular chunk will be needed (what I am seeing is ‘cat’ and not ‘cut’) after seeing the data (posterior odds) is determined by its prior probability (the base-level activation) and the amount of evidence currently available (the activation it receives from the current cognitive context). In this view each time a chunk is retrieved, its activation value is increased. However, activation is subject to decay so that the longer ago the chunk was activated, the less that activation will contribute to the current base-level activation. Then if the combined activation of a target chunk exceeds a retrieval threshold, the associated response is produced. The assumptions are also made that the latency of a response is an exponentially decreasing function of the activation level of the corresponding chunk and that the system will always retrieve the chunk with the highest activation (provided it is above the threshold). Due to the presence of noise in the system the activation values are described by a probability distribution (a logistic distribution).
Anderson et al. (1998) showed that this model can explain a number of findings from episodic memory paradigms such as serial recall, recognition memory, memory search in the Sternberg paradigm, the Tulving-Wiseman law, and free recall. For the application of ACT-R to recognition paradigms the assumption is made that during study chunks are formed in memory that encode the occurrence of the words in a given experimental list. Over the years the ACT-R framework has been extended to account for a number of phenomena related to both memory and decision making (Anderson, 2004).

Particularly relevant for the present thesis is the extension proposed by Van Maanen et al. (2012c), termed RACE/A, which incorporates an evidence accumulation mechanism within the process of memory retrieval. In RACE/A the retrieval of information from long-term declarative memory is described as a sequential sampling process that continuously accumulates evidence in favour of one of the response alternatives. More formally, the activation values of ACT-R nodes are described using an accumulation processes which piles up evidence in favour of the associated response. Notably, this model also incorporates a computational description for both short and long term dynamics of the retrieval process with the short-term component being captured by the drift rate and the long-term component being captured by the net activation.

Regarding how semantic and episodic associations are modelled this theoretical framework assumes a slightly different representation scheme than SAM-REM theories. Chunks of activation or nodes are the basic representational unit in ACT-R and are usually taken to correspond to conceptual representations. In this view semantic and episodic associations are captured by different sets of links connecting concept nodes. The influence of episodic memory is then modelled as changes in either activation or weights due to previous presentations and/or learning.

ACT-R presents an important similarity with SAM-REM theory. The net activation of a node (including the input from all connected nodes) can be seen as the ratio of activation in SAM or the odds ratio in REM. This similarity is not surprising since it is a direct consequence of the rational decision assumption. The two global models however differ in how semantic and episodic images interact. In SAM-REM the semantic and episodic memory images do not interact directly with each other but only during the retrieval process (sampling mechanism). In ACT-R instead both images are represented by the means of the very same chunk or node. The difference between semantic and
episodic relations is relegated to different sets of links connecting the nodes. This means that there is a certain amount of inter-dependence in ACT-R between the strength of activation of a semantic memory trace and of its episodic associated memories. This assumption is not enforced in SAM-REM since the activation of a semantic memory image is assumed to be independent of the activation of the episodic images associated with it. In SAM-REM the influence of semantic and episodic images occurs only in the probabilistic retrieval mechanism, which relies on both types of memory image. In ACT-R instead there is a significant amount of dependence on the activation (shared node) but not on the associations (separate sets of links). This is a subtle but important difference especially in relation to context since it is not immediately clear how a network represents the association of an item with external and internal variables (contextual feature). Since this is a point of discussion for connectionist models as well I will come back to this issue in the next paragraph.

Lastly, regarding how retrieval from memory might rely on both semantic and episodic information, an idea following RACE/A model is that the decision process cumulates evidence (drift rate of the accumulation process) from both semantic and episodic memories (intuitively the two images share the nodes but involve different types of links). In this view the activation of a single memory trace (short-term) is reflected in the rate at which the accumulator draws evidence whereas the knowledge repository (long-term) is reflected in the net activation of the system.

5.1.4 Connectionist models

The last theoretical framework I will briefly review is the connectionist approach since it shares some important similarities with classical models based on spreading activation theory (Anderson, 1998) and has produced over the last 30 years several important general models of semantic memory (McClelland, 2000; McRae, 2005).

The connectionist approach to memory, as well as to other domains of higher cognition, relies on few basic assumptions that are however qualitatively different from SAM-REM models and partly similar to ACT-R models. First, the system computing information is formed by individual neurons connected through links. Each node is a simple computing unit that integrates the incoming input from connected nodes (via sigmoid function) and when a given threshold is surpassed it propagates a signal
(produces a spike) to all of the connected neurons. The final state of activation of the system as a whole is often (but not always) taken as the end product of the computations of the system and therefore the analyses generally focus on the evaluating differences between states of activation. In comparison to previous models, this approach deviates for two main reasons. First, these models are more biologically plausible in the sense that they purposely assume a system that structurally (or functionally) resembles a biological neural network. Second, the focus of interest lies on explaining how observed effects could be obtained in a distributed parallel system meaning that computation is inherently dynamic. This focus is different from others models which instead forego the biologically constraints and rest exclusively on an algorithmic level of description.

McClelland and Rumelhart (1981) presented a relatively simple neural network model to explain a number of critical findings from word recognition. The system featured separate representations for letters and words arranged in hierarchical layers. The model was based on a simple but quite powerful principle, interactive activation. The idea was that activation while traveling back and forth between the nodes and the layers of the network would give rise to mutual interactions between different parts of the system bearing different representational content, in other words nodes at the letter levels would interact with nodes at the word level. It was shown that effects such as context effects, letter effects, and masking could all be accounted for within the framework of interactive activation theory (interactive activation model of word reading, McClelland & Rummelhart, 1981; Rummelhart & McClelland 1982).

For the present discussion of how semantic-episodic interactions can be captured within a connectionist framework the models of long-term repetition priming and long-term semantic priming are of most relevance. Authors like Masson (1995) and Becker et al. (1997) have shown first in simulation and later in behavioural work that semantic priming is not necessarily confined to the short-term dimension but can be obtained, quite robustly, also in the long-term. The classical view initially formalized as ‘spreading activation theory’ of semantic priming assumed that activation generated from a source such as a prime word would spread to the node representing an associate word and thus giving rise to facilitatory (or inhibitory) effects, priming. In the following years however it became clear that this assumption needed some revision and, as an example, ACT-R theory relaxed this aspect by assuming that activation is not confined to the spatial
dimension (spreading through links) but rather can spread by association to even spatially distant nodes.

In a distributed model of semantic memory (Masson, 1995) long-term effects of semantic priming can be obtained by means of attractor states. The architecture of the system is simple with three distinct but interconnected pools of neurons representing phonological, semantic and syntactic information (i.e. features). On each time step of computation (called a cycle), this model (as any connectionist system) presents a precise pattern of activation over all of the nodes. This pattern of activation is often called the state of the system. Throughout processing time, the system will visit many states (different patterns of activation) and the connectionist approach to memory processing is to assume that learned patterns are attractor states in the sense that the system will settle on a precise state (memory) for different initial conditions. Settling on an attractor state means the retrieval of the informational content associated with its pattern of activation. The computational principle underlying this property is pattern-completion via re-entrant connections. This principle has been extensively studied in a number of domains and it is an accepted property of both artificial and biological neural systems (O’Reilly and Munakata, 2000). The idea is simple. Through learning the weights of the connections are changed so that a particular pattern of activation is stored in the synaptic weights of the system. Later on when the system is presented again with part of the input it will autonomously settle down in the stored attractor state therefore recovering the associated memory image. Becker and colleagues (1997) have shown how in this view both short and long-term semantic priming effects can be explained, depending on whether the changes in the system are temporary (changes in activation) or more permanent (changes in weights). In this formulation long-term priming effects are due to the structuring of the synapses resulting from experience whereas short-term priming effects are due to the residual activation from previous presentations.

In addition to long-term semantic priming, repetition and morphological priming can also be explained by assuming a distributed model of semantic memory (Rueckl, 2003). The explanation for long-term repetition priming follows the same assumption as before: due to learning some states are strengthened by changes in the weights making the basis of attraction for that particular state deeper. This means that the neural network for different initial conditions will be more likely to settle down in the stored
state leading to long-term repetition priming for studied items (learnt attractor states). For the case of morphological priming however the model proposed by Rueckl (2003) presents an important difference from the initial formulation of Masson (1995), the inclusion of hidden layers between pairs of representational pools. This approach has been quite successful in explaining effects of word reading and lexical decision (Plaut, 1997; Seidenberg & McClelland, 1989). The aspect of interest here is that hidden layers provide a neural network system with the ability of going beyond their standard representational limits by building higher-order representations of the statistical regularities found in the input. Morphology can be seen in this respect as a clear example of a higher-order representational format deriving from the combination of semantic, orthographic and phonological information. It is also important to note at this point that the question of whether morphology is explicitly represented (a separate representational pool) or instead emerges due to the combination of phonology, semantic and syntax is not relevant for the present discussion. What is relevant is what features participate in these higher order representations and how are they processed by the system.

Regarding how semantic and episodic associations are modelled in the connectionist framework, the assumption is made that semantic and episodic representations are not rigidly distinct but rather entertain a relationship of gradedness between each other. Even though specific models with different flavours have been proposed, the general view is that semantic and episodic memories are specialized sub-systems participating in a highly integrated modular system (Complementary Learning Systems, McClelland, McNaughton & O'Reilly, 1995). This view imposes a graded separation between what is ascribed at the level of behaviour to semantic or episodic memory and the underlying machinery implementing the memory processes. Regarding context, it is important to note that it is not easy to define what is context and what is not in a distributed connectionist system (similar to ACT-R systems). It becomes more or less a necessity to adopt the notion of neural context and use other tools to disentangle the neural contexts relative to different dimensions of information (episodic vs. semantic). Intuitively, context is defined as the activity in the rest of the system when processing a given stimulus. Depending on whether the focus is on the semantic or episodic sub-system the neural context being evaluated will be the one more strongly associated with either semantic or episodic information. The main difference with other
more algorithmic approaches is that connectionism aims at explaining how the
processes carried out in different parts of the system (semantic and episodic systems)
combine together at the global level to yield observed behaviour (memory retrieval), for
example in the domain of consolidation (Murre, 1996) and learning (Norman & O’Reilly,
2003).

Generally speaking the connectionist approach to modelling semantic and
episodic memory is based on two key principles: 1) interactive activation constantly
drives lower and higher order computations (mutual interaction across layers) and 2)
memory is an integrated modular system with different kinds of information being
processed in localized regions (segregation) and unified information being processed at
the global level (integration). Formally, the distinction between local and global
processing in neural systems can be captured in terms of functional segregation and
integration (Tononi et al., 1998) thus providing a unified mathematical notation for all of
the reviewed models.

I will now discuss how the findings of the present thesis can or cannot be
accounted for in each of these major frameworks. The goal of this theoretical
comparison is to identify meaningful differences between these theoretical frameworks
in order to draw informative conclusions regarding the underlying computational
principles.

5.2 Model based explanations

In the previous chapters of this thesis I have shown studies that were aimed at
testing some important aspects of how semantic and episodic memory interact. In
particular I have provided evidence suggesting: 1) some features are stored in long-term
memory even in the absence of conscious awareness (Chapter 2, repetition priming of
subliminal words), 2) the rate at which information is accumulated in memory is
differentially affected by the natural frequency of a word (low vs. high) and by the type
of processing (semantic vs. episodic) (Chapter 3, cross-over interaction in drift rates), 3)
only semantically meaningful stimuli are represented in lexical-semantic memory
(Chapter 4, long-term priming for morphologically related words but not for letters),
main effect of word frequency and marginal effect of letter frequency on word
recognition (Chapter 4, normative frequency effects in lexical decision).
I will now explain in semi-quantitative terms how these four main findings can (or cannot) be accommodated in each of the theoretical frameworks (global models of memory) previously introduced.

5.2.1 Subliminal repetition priming for words (Chapter 2)

In the second chapter I presented results from a series of experiments aimed at testing the hypothesis that words which have been processed subliminally (without the focus of attention and without episodic awareness) lead to the emergence of repetition priming effects. In other words, whether visual awareness is a requisite for the storage of information in long-term memory. The results indicated the presence of a small but rather robust priming effect for subliminal words marked by a slight increase over repetitions in accuracy performance and constant reaction times. As argued in the discussion section of Chapter 2, a plausible computational explanation can be formulated in the framework of SAM-REM theory. The same mechanisms that have been suggested to account for implicit priming phenomena in recognition and recall (REM-I) on the one hand, and lexical decision tasks (REM-LD), on the other hand, may also account for the observed subliminal long-term priming effects. The crux of the explanation rests on two points. First, during subliminal presentations contextual features are added not only to the episodic memory trace but also to the semantic memory trace. In SAM-REM theory this could be justified by assuming that when the images are sufficiently similar, contextual features are also added to the lexical-semantic trace. Following this explanation, the encoded contextual features from the study episode could favour later on the memory decision process by providing additional matching features, thus leading to a facilitatory effect in behavioural performance (repetition priming). Crucially, when the pairing of prime-targets words was varied (Experiment 6) no differences were found between fixed and recombined pairs. Overall these findings seem to suggest that the subliminal priming effect was mainly driven by context features rather than item specific or content features.

This interpretation is also in line with some models of visual processing and masking (Enns & Di Lollo, 2000). According to the model of visual masking by object substitution the processing of the mask during encoding has the effect for subliminal items of disrupting the binding of features active in working memory with the
represented objects. Following this interpretation the prediction for masked items is that only features pertaining to the whole study event (contextual features) are stored in long-term memory. This explanation is consistent with the assumption in SAM that the retrieval in tests such as perceptual identification starts with a cue that includes both contextual and perceptual features. In this view context favours the sampling of those images that are more strongly related to the probe features. In other words the semantic image that presents the highest number of matching features with the study episode (the presented prime) will tend to be retrieved. A similar and equivalent account could also be formulated in REM theory by assuming that during masking there is storage of noisy features (encoding more error prone) that ultimately leads to less matching (and more mismatching) features for subliminal items.

In ACT-R subliminal effects could be explained by assuming a lower threshold. In the RACE/A extension for example this could translate to a higher net activation value for subliminal items compared to novel items and to lower drift rates reflecting the fact that subliminal priming is driven mainly by contextual information and not by item specific information.

The interactive activation model proposed by McClelland & Rummelheart (1981) for word recognition has been shown to account for contextual effects of letters embedded into words during masked presentations. In this model the repetition priming for subliminal words is explained by the interactive activation generated by other items that share the same context (list). In this view it is the combined activation that drives the system to select the item with the largest number of matching features (in a fashion similar to SAM’s probabilistic retrieval). Also in this case the assumption needs to be made that during masked presentations the encoding of features is noisy and more error-prone in order to predict a slowly increasing effect of repetitions and no effect of prime-target pairing. It is interesting to draw an analogy here with the fact that connectionist systems have been shown to be incapable of learning multiple overlapping patterns unless these patterns are learned in an interleaved fashion (French, 1992; McClelland et al., 1995). In this respect the failure of binding features with represented objects could be seen as the result of interleaved learning ‘cancelling out’ information pertaining to specific items and preserving (holistically) information pertaining to the whole study episode.
5.2.2 Differential drifts of word frequency over memory tasks (Chapter 3)

In the third chapter I focused on a slightly different question, namely whether behavioural performance in lexical decision and recognition, two laboratory tasks commonly associated with semantic and episodic forms of memory processing, respectively, are based on different processing components. One of the central findings of this study was the fact that behaviour across both tasks was explained by a theoretical model of memory sensitive to two dimensions and, more importantly, that these dimensions were found to entail different processing components (i.e. different drift rates). The link between the multi-dimensionality of memory and the processing components deployed was established by assuming that information cumulates in the underlying memory system as an accumulation process (as in RACE/A model). Following this view we estimated the memory components associated with each dimension (task) by means of the linear ballistic accumulator model (LBA). The crucial result of this study was that the rate at which information is accumulated in memory differs significantly between semantic and episodic dimensions with an inversion of the effect of normative word frequency over task-set. In episodic processing low-frequency words led to a higher rate of accumulation (drift rate) compared to high-frequency words and this pattern completely reversed for semantic processing with an advantage of high frequency words over low frequency ones. Overall these results suggest that a similar mechanism is deployed for the retrieval of information from long-term memory (information accumulation) across tasks. What changes is the quality of the information extracted from memory and therefore the rate of evidence accumulation for the decision at hand. Intuitively this could suggest that the same features (i.e. low-frequency) that are highly diagnostic in one condition (episodic processing) are instead only marginally diagnostic in the other condition (semantic processing). In RACE/A model it becomes natural to account for this inversion in drift rate by assuming separate accumulators for episodic and semantic processing.

In SAM-REM theory the difference in rate of accumulation could be translated to either changes in the ratio of activations of the memory images (SAM) or to differences in the posterior odds between the matching and mismatching features (REM). As shown when introducing SAM and REM global models, these two quantities are mathematically equivalent. The point of interest here is how any of them could be translated to
information accumulation. The RACE/A model provides an example by relating the rate at which evidence cumulates with the activation generated in the nodes of the declarative system. Following this approach in SAM-REM theory the explanation would be that a higher drift rate corresponds to a ratio of activation or posterior probabilities in favour of the numerator (target image). The core assumption underlying SAM, REM and ACT-R frameworks which elicits this interpretation is the fact that the decision process is rational and thus maximizes the amount of evidence provided by both matching and mismatching features, regardless of this representational format.

In the connectionist framework a model along the lines proposed by Masson and Becker could account for this finding. The linear ballistic accumulator (LBA) model can be seen as a natural extension of the connectionist view since the time course of a decision process can described also as a leaky accumulator process (Usher & McClelland, 2001). A first notable difference between these two models is that in one case the process cumulates information linearly and without decay (LBA) whereas in the other the process is non-linear and information decays over time (leaky integrator). A distributed model of semantic memory that also implements a module for episodic memory can be seen as a competing leaky integrator. In such a system the inversion of drift rate could then be explained by the fact that in one condition the semantic module drives the decision process (higher drift rate) whereas in the other condition this reverses with an advantage of the episodic module.

5.2.3 Morphological priming (Chapter 4)

In the fourth chapter of the present thesis I have investigated two specific questions regarding the internal function organization of semantic memory: 1) whether the storage of episodic features in long-term memory is confined exclusively to meaningful units and 2) whether manipulating normative frequency for words and letters yields distinct effects in lexical decision.

The presented experiments showed two important results. First, I found evidence for morphological priming indicating that semantic memory encodes higher-order representations (morphology) but not perceptual ones (letters). The explanation proposed for morphological priming was that features are updated in long-term memory only for semantically meaningful units but not for perceptual units, which in
many theories is the hallmark of semantic memory (representations with meaning). In this view letters, which do not bear a representation in semantic memory, are not expected to generate priming effects. In the first experiment of Chapter 4 we found a priming effect for words and morphemes but not for letters supporting this interpretation.

In SAM-REM theory morphological and word priming would be a direct consequence of the fact that semantic memory does not possess representations of perceptually meaningful objects, letters, and therefore there cannot be any updating of information in long-term memory thus no long-term repetition priming.

In the ACT-R framework instead morphological priming could be explained by assuming some form of hierarchical structure representing different lexical items (morphemes and words) embedded in the connectivity (links) of the nodes of the declarative system. The long-term repetition priming effect then would be a result of the combination of different conceptual nodes and of the modified weights linking them. Similarly to the explanation of SAM-REM theory, the assumption would be made that letters elicit at most temporary activation whereas meaningful units trigger weight adaptation.

Connectionist models have provided some important insights regarding the nature of morphological priming (Rueckl, 2003). In this model each representation is a pattern of activation spanning the three neuronal pools for syntax, semantics and phonology with the assumption of hidden layers between representational pools (Ruckel et al., 1997). On the one hand, the interactive activation that flows through the system gives rise to representations of higher-order. Morphology in this view results from the hierarchical combination of phonological, orthographical and semantic information. On the other hand, processing the study items leads to a change in the weights connecting the nodes (the knowledge of the system) that promotes later on at retrieval an advantage for the previously studied item (deepening of attractors). It is important to note that even though an explicit representational pool could be included for morphology, it is doubtful where that would be appropriate. If the memory systems were to encode explicitly every type of hierarchical information they need to process, there would be an explosion of resources required. A more parsimonious approach following Occam’s razor is to represent explicitly only those dimensions that are the
principal components of the representational space and to abstract hierarchical
representations for other domains by combining these fundamental components.

In comparison with SAM-REM theory it is worthwhile to highlight another
difference. SAM-REM theory does not specify how morphemes are represented;
however an intuitive idea is that the updating of long-term memory with information
(i.e. features) proceeds in the same fashion regardless of whether the memory image
represents a word or a morpheme. The reason for this could be, for example, that the
morphemes activate the most similar noun stored in semantic memory and thus lead to
a similar updating in long-term memory. In the connectionist view instead, since there is
no explicit representational pool for morphology, the updating of information has to be
found in changes in the attractor dynamics of the system. The repetition priming effect
for morphology therefore is to be found in this case in the change of the orbits
(deepening) around the attractor states of specific morphemes previously studied
(Rueckl et al., 1997). The interesting aspect of this comparison is that both frameworks
can equally well account, in principle, for the presence of repetition priming for word
and morphemes but not for letters. This seems to suggest that the relevant aspect to
explain this priming pattern is not whether morphology is explicitly or implicitly
represented in memory, but whether there is some form of hierarchical structure in its
representation.

5.2.4 Word and letter frequency effects (Chapter 4)

In the second experiment of Chapter 4 I looked at how manipulations of
normative frequency, that is the number of times a word or a letter occurs in written
text, affects lexical performance and thus the retrieval of information from long-term
semantic memory. The crucial finding of this study was twofold. On the one hand,
normative word frequency led to faster and more accurate responses for high frequency
words compared to low frequency ones. On the other hand, normative letter frequency
showed no effects in accuracy and led to faster responses only for low frequency words
with a slight advantage of high frequency letters over low frequency ones. Over
repetitions subjects’ responses became increasingly faster for both high and low
frequency words and more accurate for low frequency words. Repetitions showed an
interaction effect with word frequency but not with letter frequency.
In line with previous suggestions from the literature arguing against a distinction between memory processes in terms of ‘pure’ task performance (McKoon et al., 1986; Ratcliff et al., 1985), the explanation proposed is that this pattern of results is due to the contribution of both semantic and episodic memory during lexical processing. The interaction between normative word frequency and repetitions, in particular, can be seen as an interaction between semantic and episodic memory systems.

In SAM-REM theory the word frequency priming effect could be naturally accounted for by the fact that high frequency words present an increase number of matching features with the lexico-semantic memory trace in comparison with low frequency items. The repetition priming effect could be explained by assuming, as for the discussion of Chapter 2 results, that the lexico-semantic trace is updated with episodic (contextual) features from previous presentations. This could also provide, in principle, an explanation for the interaction between word frequency and repetitions; for example by postulating that more features are updated for the memory images of low frequency words since the image itself is more incomplete. The effect of normative letter frequency in speed of processing could be explained similarly in SAM-REM. Since memory images for low frequency words are assumed to be more incomplete, they will have less matching features with the probe and thus a lower diagnostic value for the decision process. From this it follows that the lower the normative word frequency is, the more informative normative letter frequency is for the lexical decision. However, the fact that normative letter frequency had an effect only in terms of speed of processing (i.e. reaction times) and not in terms of quality of processing (i.e. proportion correct), seems to suggest that letter frequency effects are more related to perceptual features whereas word frequency effects are more related to content features. In sum the explanation of the results proposed in SAM-REM is that normative word frequency effects are due to different amounts of diagnostic evidence for the sampling process whereas the repetition priming effect is due to the updating of the lexico-semantic trace with episodic (contextual) features from previous presentations.

In ACT-R the effect of normative word frequency could be explained similarly to SAM-REM by assuming a higher activation value for nodes that represent high frequency words in comparison to nodes that represent low frequency words. This increased activation for high frequency words could be explained similarly to SAM-REM by assuming more activation at the feature level (i.e. letters) due to, for example, an
increased number of matching features with the probe. The word frequency effect could be explained also in this case by a higher diagnostic value for high frequency words in comparison to low frequency ones (rational decision process), reflected in this case in an increased level of activation for nodes representing high frequency words (over the activity of the whole system). The effect of word frequency could be ascribed then to the links connecting the nodes with higher weights (more activation spread) for links connecting higher frequency words. It is important to note that as in SAM-REM, the same idea could be applied to explain why letter frequency is diagnostic only when word frequency is low. In this case the explanation would be that lower frequency words generate overall less activation (due to lower weights). Therefore the activity generated by the feature level (where letters are represented) can have a higher impact on the activity at the word level, therefore having more diagnostic value for the decision process; but only when the overall activity is low (low frequency word). A classical explanation of the repetition priming effect instead could be in terms of baseline activity of each node. With each presentation the activation value of the corresponding node is increased, therefore leading to an accumulation of activation in the baseline value of the node. This selective increase in baseline activity across multiple presentations can predict the repetition priming effect.

In connectionist systems instead an explanation could be formulated within the interactive activation model of McClelland & Rumelhart (1981). It was shown that this model could account for letter frequency effects in relation to word frequency. The explanation is based on the interactive activation principle and proposes that, on the one hand, low level units representing letters (letter units) send activation to higher level areas where words are represented (word units) and, on the other hand, activation from higher level areas is send back to lower levels. Analogous to the case of ACT-R, the main effect of word frequency effect and the marginal effect of letter frequency could be explained by assuming higher weights for high frequency words (or letters) and lower weights for low frequency words (or letters), with overall smaller weights for the letter units in comparison to the weights for the word units. In this scenario the activation originating from low-level nodes (letter units) would be smaller in comparison to the activation generated at the higher-level (word units), therefore the activity generated by letter units would have an impact only when the activity generated by word units is
small enough. Repetition priming as well could be accounted for by assuming changes in the baseline activation values of associated units with each stimulus presentation.

5.2.5 Derived assumptions

To formally characterize the systematic interactions between semantic and episodic memory as they emerge in behaviour a minimal set of three models is most likely needed: a model for episodic memory, a model for semantic memory and a model for their interaction. The inclusion of an explicit model of the interactions between memory systems is dictated by the fact that the behavioural performance measured experimentally represents the final output of a system that is inherently multi-dimensional. In the present context multi-dimensional means that the underlying source of information is not unitary but rather composed of multiple principal components.

Across multiple experiments I have looked at some core aspects of the interaction between semantic and episodic memory and in particular at how this can change across conditions (or tasks), leading to systematic differences in the retrieval mechanisms and thus in behavioural performance. A model of memory interactions is therefore needed to define the systematic relationships encompassing the underlying memory resources (i.e. semantic and episodic dimensions) and the behavioural tasks used to measure performance (i.e. processing components). Put it differently, a model is needed to define the mapping between (multi-dimensional) memory strengths and behavioural performance.

From the general discussion of the previous section on how different theoretical frameworks can in principle account for the observed effects, a few computational principles emerge as a common denominator across them. What follows is a short summary of these computational principles alongside some final considerations of their relevance for a model of semantic-episodic memory interactions.

The first assumption is that retrieval of information from long-term memory is a probabilistic process and it is a general core property of all the global models of memory discussed previously. The second assumption is that the memory decision processes evaluates information in a rational or Bayesian fashion as explicitly modelled in SAM-REM and ACT-R frameworks. In the connectionist models reviewed earlier this principle was not discussed since the type of networks implementing such principle (Dynamic
Bayesian Networks, Rao, 2004) constitute a deviation from the classic neural networks used in the other reviewed models. It suffices for the present discussion to point out that a classic connectionist system can be seen as a rational decision maker when a few qualitatively different assumptions are made (for a review see Murphy, 2002). The third assumption is that behavioural memory performance can be supported by both semantic and episodic memory systems in parallel with mutual interactions criss-crossing the two systems. Critically, since these interactions can change depending on the environmental demands, a model of their interactions is needed to specify how the two memory modules are configured as parts of the same large-scale modular system; in other words a description is needed of how the dimensions of memory combine to generate behavioural performance.

In summary the proposed interpretation is that during long-term memory retrieval the interactions taking place between semantic and episodic memory dimensions are highly dynamic but with a significant degree of systematicity. From a preliminary discussion of the present results in light of current models of memory, the interactions between semantic and episodic memory modules observed during LTM retrieval seem to confirm with standard generalized principles: a probabilistic mechanism for retrieval from LTM combined and a rational decision process for evaluating available evidence against information stored in memory images. In virtue of the overlapping machinery implementing both systems (shared processing components or neural networks), there is some evidence that these interactions should not be regarded as a fixed relation but rather as a global property that can change as a function of the environmental (and local) demands.

In the next concluding section I will highlight some suggestive findings from the literature that point to a similar interpretation in a different cognitive domain.

5.3 Final considerations

The analogy I would like to suggest in this concluding section follows from what has been proposed in the literature to explain how different brain rhythms are produced in relation to motor planning (Rabinovich, Varona, Selverston & Abarbanel, 2006). The study of how different functional organizations can emerge from the same structural system has been one of the major challenges for cognitive neuroscience (Sporns &
Kötter, 2004). In the motor domain, a number of studies have been conducted in an attempt to elucidate how different neural structures could be functionally organized during behaviour to produce a coherent rhythm and what relevance they could have for complex forms of cognition; the hallmark of their functional organization being the specific brain rhythms produced during a particular behaviour. A stimulating view that has been proposed to characterise this aspect of brain networks, how the same structural configuration can give rise to different functional configurations, is in terms of a liquid state machine. A liquid state machine is a mathematical definition introduced in dynamical systems theory to describe (and understand) a number of multi-dimensional systems. Multi-dimensional systems are complex systems where the interactions between each of its constituent parts can take place at different levels of description and at different time scales. This concept can be appreciated with an intuitive example. Take a bucket of cold water and assume that a block of ice is floating on top. Both the water and the ice are composed of the same chemical compounds (H\textsubscript{2}O) however there is a difference in their current state. Water is in a liquid state whereas ice is in a solid state; therefore the two systems share the same structural configuration (H\textsubscript{2}O) but different functional configurations (liquid-solid). The crucial point of this analogy is that each functional configuration endows specific properties to the system (liquid vs. solid) and the two are in constant mutual interaction due to the surface where they touch (the ice floats over the water).

The brain is a clear example of a multi-dimensional system and there is an increasing interest for viewing cognition as a form of computation that relies on transients more than on attractor states (see Mass, Natschläger, Markram, 2002; Büsing, Schrauwen & Legenstein, 2010). The characterization I would like to propose in conclusion is that long-term declarative memory can also be seen as a liquid state machine where the semantic and episodic dimensions of memory can be crystalized in specific configurations to produce particular cognitive activities such as a memory for events and a memory for knowledge.