1
Paradigms, Principles, and Processes

A brief history

The study of human memory has a long and rich history spanning the wisdom literatures of ancient civilizations and continuing through the philosophical tradition of the British empiricist and associationist movements (Yates, 1966). The scientific study of human memory began in Germany in the late 19th century with the classic studies of Hermann Ebbinghaus (1850-1909) and George Elias Müller (1850-1934). These early studies demonstrated the power of experimental design and quantitative measurement in the study of memory, and introduced the classic "list learning" methodology, which has provided the basic experimental framework for the study of memory until this day (Ebbinghaus, 1885/1913; Müller & Pilzecker, 1900). Scientific research on human learning and memory quickly spread to the United States, where new laboratories sprung up at Harvard, Princeton, Cornell and Clark Universities. As in any new science, the early work was largely descriptive, aiming to characterize and quantify the basic data obtained from different experimental techniques (S. E. Newman, 1987).

Subsequent decades saw the emergence of several prominent "schools of psychology," each taking a different theoretical approach to the analysis of memory. The Gestalt school emphasized the role of mental representation and the organization (and reorganization) of knowledge as critical aspects of learning and memory (Köhler, 1947; Katona, 1951), whereas the associationist school emphasized the power of experience to create new webs of associations among pre-existing representations (Robinson, 1932; McGeoch, 1942). Influenced by the study of animal learning, many theorists emphasized the importance of reward-seeking behavior as the driver of learning and downplayed the idea that memory was anything more than a means of enabling animals to obtain rewards through action (Keller & Schoenfeld, 1950; McGeoch & Irion, 1952).

In the first two decades following World-War II, information theory and the rise of computing machines provided a powerful new framework for thinking about memory (and cognition more generally) as a result of sequential stages of information processing. Here the idea is that the brain is a computing device that transduces sensory inputs into patterns of electrical activity and processes these patterns through sequential but possibly overlapping stages with the goal of achieving some behavioral objective. This new framework eventually transformed the field, sparking a "cognitive rev-

olution," in which theorists began to think of memory as one central part of a computing machine whose "software" could be decoded through analyses of accuracy and response time in well controlled experiments. Scientists could now peer into the "black box" (a term used to describe a computing machine whose inner mechanisms were beyond analysis) and hypothesize the underlying machinery supporting memory and cognition.

In parallel with the aforementioned developments in the cognitive analysis of human memory, scientists studying the biological mechanisms of the brain in animals developed techniques for measuring electrical signals in the brain while animals were engaged in complex behaviors. These developments, which will be further discussed in Chapter 2, set the stage for new theories of memory based on the idea that nerve cells (rather than microprocessors) were doing the computing operations supporting memory and cognition. This, in turn, led to the emergence of "connectionist" or "neural network" models of memory and learning in the 1980s, and the subsequent emergence of the field of computational neuroscience in the 1990s.

The late 20th century saw the emergence of the personal computer. This technological advance had transformative effects on both the experimental and theoretical analysis of human memory. Personal computers quickly filled psychology laboratories around the world, enabling students to rapidly design and deploy experiments that precisely controlled the presentation of lists of words or pictures and allowed researchers to easily collect a rich array of data on behavior, including the measurement of response time, mouse clicks, etc. As the graphics capabilities and processing power of the computers improved, researchers began developing new types of experiments, including those based on video-game technology (desktop virtual reality) that enabled the study of spatial memory within realistic 3D-rendered environments. Computing power also helped to fuel the development of computational models designed to account for the rich experimental data obtained in these laboratory experiments. With the rise of the internet in recent decades, researchers have now uploaded their experimental laboratory to the "cloud," enabling simultaneous data collection from large numbers of people as they perform game-like cognitive tasks controlled within their web browsers. This, in turn, has allowed scientists studying memory to examine behaviors of people who are not university students, thus greatly expanding the applicability of their research to the world's diverse population.

The focus of the present work is on the electrophysiology of human memory, which has been part of the remarkable transformation in our understanding of human memory emerging from cognitive neuroscience methods in recent decades. Whereas memory research during the 20th century primarily advanced through the analysis of overt behavior (either accuracy, response time, or both), 21st century research is accelerating through the use of technologies that allow us to peer into the brain as humans study and retrieve information. This technology includes both measures of bloodflow as measured non-invasively using functional-magnetic resonance imaging, and measures of electrical activity that can be obtained either non-invasively using scalp recorded electromagnetic signals (EEG and MEG) and invasively by studying patients who have electrodes implanted in their brain for the treatment of various disorders. Both invasive and non-invasive electrical stimulation of the brain is also providing a powerful tool for the analysis of learning, memory and a variety of perceptual and cognitive functions. These

recording techniques have led to advances in our theoretical understanding of how humans learn and remember information, and they are also setting the stage for new therapies that may help treat disorders of the brain that affect human cognition.

Laboratory paradigms

The standard paradigm for the study of human memory, whether using sophisticated brain recording techniques, or internet-based memory games, still involves the presentation of sequences of to-be-remembered items (*memoranda*) which research participants (*subjects*) are asked to *recognize* or *recall*. Here we briefly describe the basic methodology used in recognition and recall experiments. In later sections we will introduce important variants of these methods as well as some other techniques.

Recognition

Consider a police lineup in which the witness to a crime is later faced with the task of judging whether a suspect in the crime is the actual perpetrator. This is an example of a recognition task. Someone has experienced some information (e.g., having seen a person commit a crime) and must later decide whether a test item (the appearance of a suspect in the lineup) matches the memory of the person seen committing the crime. Here, the judicial system would very much like to understand what factors influence witness accuracy and whether the circumstances of the lineup can be engineered to improve witness accuracy (e.g., Coloff and Wixted, 2019).

Recognition judgments are commonplace. In watching a rerun of a favorite television program, you may ask yourself whether you have seen this episode before. If the answer is yes, your level of interest may be reduced (unless you remember it as being particularly good). If you forget to bring your shopping list to the store, you may walk the aisles trying to recognize particular items as being among those you planned to purchase. Daily life contains numerous such circumstances that call upon our ability to make recognition judgments.

In studying recognition memory, scientists would like to exert greater control over the conditions prevailing both when an item is first studied and when memory for the item is later tested. Thus, in a laboratory recognition task, participants study a list of items presented in a highly controlled manner. These items are usually words, but other nonlinguistic stimuli such as complex scenes may also be used. During the test phase of the experiment, participants are shown a mixture of items that were presented as part of the study list and items that were not presented as part of the study list. For each item they must indicate whether or not it was on the just-studied list (Figure 1.1 illustrates the basic recognition procedure).

Recognition-memory judgments can be made for any attribute of an item, not just its membership on the list. For example, one might ask people to judge which of two items was most recently presented, how many times a given item was presented, or whether an item was presented in a male or female voice. Recognition can also be tested in a multiple-choice format. In this case, two (or more) test items are shown and participants have to judge which appeared on the list. Rather than asking subjects to make binary

choice decisions, we can ask subjects to make ratings of confidence, or even to rank a set of choices based on their likelihood of having been studied on a list.

Recall

Recall is a more active form of memory in which we actually remember information that is distinct from the test cue itself. In taking an essay exam, a student must recall previously studied information that is relevant to the specific question (and hopefully organize that information in a coherent and insightful manner). In meeting an old friend, one would hope to be able to recall their name, when you last met, and other shared experiences. We have all experienced the frustration of trying to recall something that we feel we should remember, but at a given occasion we are unable to do so (and then, when it is hardly needed, the information easily pops into mind!).

In the laboratory, we can study recall in a variety of different ways. After studying a list of words, participants may be asked to recall the items freely, in any order (*free recall*), in order of study (*serial recall*), or even backwards. Figure 1.1 illustrates the free-recall task. Participants may also be probed to recall specific items. For example, the fifth word in the list may be given as a cue, and the participant may be asked to recall the prior or the subsequent item (*probed recall*).

One of the most widely used recall tasks involves the study of paired items followed by a memory test in which one member of each pair is shown as a cue for recall of its mate. As an example, you might study randomly paired common nouns, such as *carriage–dog*, *tree–pen*, *ribbon–diamond*, *horse–house*, etc. At the time of the test, you might first be asked to recall the word paired with *pen*, then the word paired with *ribbon*, and so on. This task is alternatively called *paired associates* or *cued recall*. Pairs can be probed either in the forward direction (e.g., *ribbon* as a cue to recall *diamond*) or in the backward direction (e.g., *pen* as a cue to recall *tree*). Although one might expect forward recall to be easier than backward recall, research has shown that people's ability to recall an individual pair does not depend on the order in which the pair was studied (Kahana, 2002). Yet, despite this fact, people can remember which item was first and which item was second (Mandler, Rabinowitz, & Simon, 1981).

Although words are the most commonly used stimuli in recognition and recall tasks, one can also use nonlinguistic materials, such as abstract pictures or shapes. One reason why memory researchers have emphasized the use of linguistic materials is that adults are very likely to attempt to code confusable nonlinguistic materials, such as complex shapes, in a linguistic manner. In the case of recall tasks, it is also difficult to measure the recall (reproduction) of nonverbal materials such as pictures. On the other hand, perceptual imagery plays a critical role in the encoding of linguistic materials, and visual imagery can greatly facilitate many forms of learning (Paivio, 1986).

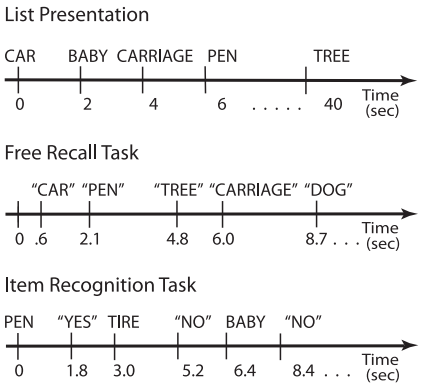

Figure 1.1: Recall and recognition tasks. Top: timing of item presentation. Middle: sample responses during a free-recall task. Bottom: sequence of test probes and *yes-no* responses in an item-recognition test. The words *pen* and *baby* appeared on the studied list, whereas the word *tire* had not. Quotes denote participants' responses.

Other paradigms

Although recognition, recall, and their variants are the methods most commonly used in the study of human memory, there are other important classes

of memory paradigms that we will briefly introduce. Recognition and recall are often referred to as direct probes of memory because the subject is trying to consult their memory for the prior occurrence of an item. In other measures, memory may be proved indirectly, for example by having subjects perform a judgment task on an item that was previously encountered. For example, in a *lexical decision* task subjects must decide, as quickly as possible, whether a sequence of letters is a valid word. Participants rarely make errors at this task, but they will respond more quickly to words that were recently encountered. In a *perceptual identification* task, participants see a degraded version of a word or picture that they have to identify. Again, identification is usually faster and more accurate if the item has recently been encountered, though there are important exceptions to this general observation (Huber, 2008). The direct–indirect distinction refers to the way in which memories are retrieved. In direct memory tests, participants are asked to recognize or recall an item as having occurred in a list of previously studied items. Indirect memory tests do not make reference to a previous study episode. Rather, they reveal the memorial consequences of the study episode on a person's ability to perform some other cognitive task, such as perceptual identification or fragment completion.

A related distinction applies to the study phase of a memory experiment. When presenting participants with a set of material for study, the experimenter may instruct them to *intentionally* learn the materials for a subsequent memory test. Alternatively, the material may be presented *incidentally* by using some cover story designed to make it less likely that participants will expect a memory test. For example, participants may be asked to rate the pleasantness of each word in a series or they may be asked to judge whether the words correspond to living or nonliving things. Although intentional learning is most often studied in the lab, incidental learning may be more typical of the operation of memory in our daily lives, where we experience a variety of information without necessarily trying to learn it for an upcoming test. Even if you are not trying to remember a specific detail of an experience, you may nonetheless store that detail in memory, and later, it may pop into your mind without any specific intention to retrieve it.

Principles

Encoding and Retrieval

Most memory tasks involve separate study and test phases. During the study phase, subjects *encode* new information into memory and during the test phase they *retrieve* the previously learned information. Researchers generally use the terms encoding and retrieval to refer to memory operations or processes, but sometimes for convenience they also use these terms to refer to the study and test phases of an experiment. However, this is not quite correct. As you can well imagine, while studying new information, subjects are also likely to retrieve prior knowledge about that information from memory and associate that prior knowledge with the new learning context. Similarly, when retrieving information from memory, subjects are also likely to be encoding the retrieved information into the test context. Indeed, research shows that the best way to learn something is to practice retrieving it (Karpicke & Roediger, 2008; Roediger, Putnam, & Smith, 2011).

Five laws of memory

Although many psychological phenomena appear to be fickle, being observed only under highly specific conditions, there exist a set of phenomena concerning memory that appear to be quite general, even universal. These phenomena appear across a very broad range of conditions, in virtually all people, and even in a variety of non-human animal species. Here I refer to these general principles of memory as "laws of memory." Following Teigen (2002) and Roediger (2008) I use the word law simply to refer to an empirical regularity that appears in a very broad, nearly universal, set of conditions. I recognize that for any law we might identify there may be bounds on its range of applicability, much like the celebrated laws of physics may be experience violations, for example, at extreme values of speed or size. Here we review five major laws of human memory: the law of recency, the law of primacy, the law of contiguity, the law of similarity, and the law of repetition. I suggest that these phenomena may be considered laws of memory, with the implication that a serious theory of memory should make predictions about these phenomena that appear consistent with the extant data. These laws of memory should also relate to neural mechanisms that would likely appear in measures of the brain's electrical activity, as discussed in subsequent chapters.

The Law of Recency

The *Law of Recency* canonizes the observation that recent memories are more easily retrieved than memories of more remote experiences (Brown, 1824; Calkins, 1896). In his early studies, Ebbinghaus demonstrated that forgetting is rapid at first and then gradually slows. Figure 1.2 illustrates the forgetting function in three classic memory paradigms: paired-associate memory, free recall, and item recognition. In each of these three studies, forgetting is illustrated over the course of an experimental session, ranging from 10 minutes (A), to 40 minutes (B) to 70 minutes (C). In each case, we can see that forgetting continues over the course of the retention interval. In other experiments, people have studied forgetting effects over the course of many years, but under those circumstances it is much harder to control for the mental activity of the subject during the retention interval.

This seemingly obvious aspect of memory has generated more research and controversy than perhaps any other. One of the most important early findings was that the rate of forgetting depends critically on what happens between initial study and later memory tests (Kahana, 2012). Forgetting tends to be greatest when the interval between study and test (the *retention interval*) is filled with high levels of mental activity. In most experiments, forgetting of a given word pair is rapid because the retention interval is filled with studying and getting tested on other word pairs. In a classic study of forgetting over the course of sleep or waking activity, Jenkins and Dallenbach (1924) showed that subjects forget far more quickly when the retention interval is filled with waking activity than when subjects are asleep. Modern neuroscientific studies have focused considerable attention on the role of sleep in memory and an active debate currently centers on the question of whether sleep-related processes may actively help in the consolidation of memories learned during the day.

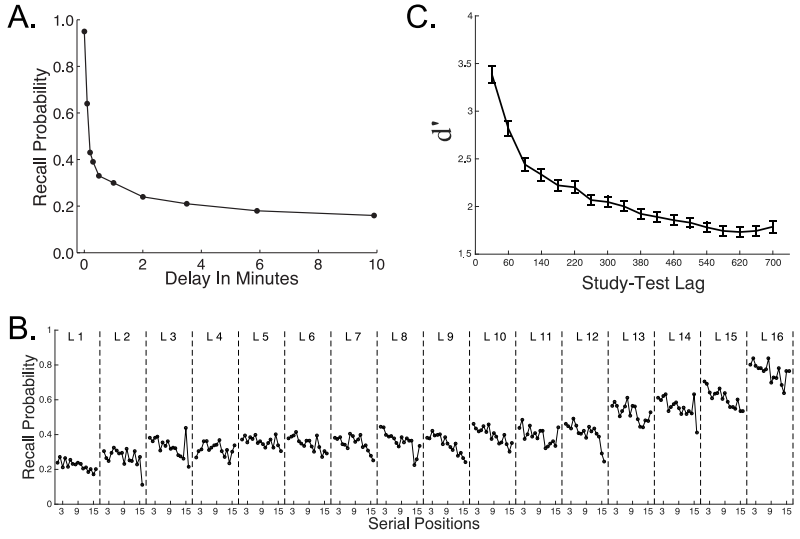

Figure 1.2: Forgetting in three classic memory paradigms. A. Memory for paired associates as a function of the amount of time (and number of pairs) intervening between study and test. Data from Rubin et al. (1999). **B.** Recall probability in final free recall as a function of both serial position and list position. More recent items appear towards the right. **C.** Recency in item recognition with performance measured using d', which is the z-normalized hit rate minus the z-normalized false alarm rate (Data from Experiment 1 of the PEERS study).

Short-term and long-term recency

The above figures illustrate recency on the scale of a single hour-long experimental session. Recency is also seen on a much faster scale by looking at serial position effects in recall of items within a single list. Here we find that the forgetting is very rapid. In the case of studying a list of 12 novel word-word pairs (paired-associate memory) if the last pair is tested immediately following study, it is almost always recalled correctly, and the next-to-last pair is recalled more than 50% of the time. Earlier pairs, however, are recalled less well, and the level of recall of these pairs decreases very slowly with increasing study-test retention intervals (Murdock, 1967). This can be seen for data from two different experiments, as shown in Figure 1.3.

The modality effect. The recency effect for paired associates is larger for auditorally presented lists than for lists that are presented visually. This *modality effect* is usually limited to the last 2 or 3 studied pairs (Murdock, 1972). The recency effect is also larger for auditorally presented items in other recall tasks, such as free and serial recall.

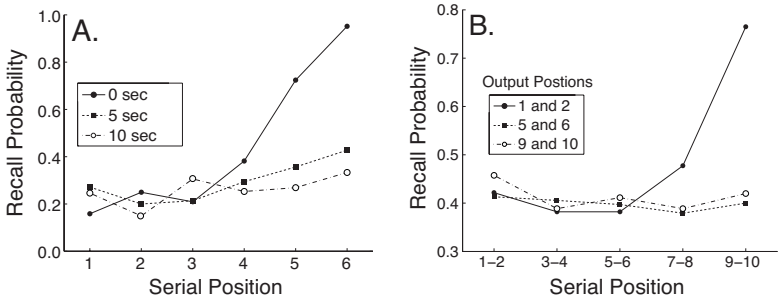

Figure 1.3: Short-term recency. People are very good at remembering the last couple of pairs in a list when they are tested immediately after study. **A.** Recency is greatly reduced after a delay of 5 sec (Murdock, 1967). **B.** Forgetting occurs during the recall phase itself. Items tested early in recall (output positions 1 and 2) show marked recency, whereas items tested late in recall (output positions 9 and 10) do not exhibit recency (Tulving & Arbuckle, 1963).

The Law of Primacy

When learning a series of items in a new context, subjects often exhibit superior memory for the first item in the series, and this advantage often extends

through to several early list items. We can readily observe this *primacy effect* in the free recall task, where subjects may recall the list items in any order. This is shown in figure for data obtained in the study of free recall. Subjects studied lists of 15 common nouns, which they subsequently recalled after a period of distraction. Here one can see a marked primacy effect in which recall is highest for the first list item and falls to a stable baseline by the fourth or fifth list item.

The classic explanation for primacy is that subjects, who are highly motivated to recall as many items as possible, share rehearsal time between the currently presented item and items presented in earlier serial positions (Brodie & Murdock, 1977; Tan & Ward, 2000). Experimental manipulations that tend to promote rehearsal, such as providing subjects with longer interpresentation intervals, also enhance primacy. By contrast, conditions that discourage rehearsal, such as incidental encoding, fast presentation rates or interitem distraction, reduce primacy effects (Kahana, 2012).

The primacy effect is by no means specific to the free recall task. Primacy effects appear in delayed item recognition for both short and long lists. Paired-associate and probe-recall tasks also exhibit primacy effects, although the effects are often of smaller magnitude. Primacy effects also appear following event boundaries in both word lists and more naturalistic experiences As an example, (Polyn, Norman, & Kahana, 2009) asked subjects to study a list of twelve items in which they judged each item's size ("big" or "small") or pleasantness ("good" or "bad") . In some lists, subjects were asked to make the same judgement on every item. In others, subjects either made size judgements on the first six items and pleasantness judgements on the last six items, or vice versa. The lists in which subjects had to shift tasks midway through exhibited two primacy effects: a primacy effect for the beginning of the list, and an additional primacy effect after the task shift boundary. Thus, a change in the manner of item encoding resulted in improved memory for the first few items studied under the new encoding conditions.

While primacy effects exhibit near-universality across memory paradigms and subjects [1] , scientists are still debating the exact mechanisms underlying this phenomenon Although rehearsal of early list items constitutes an important factor underlying primacy in the free recall task, this mechanism cannot explain primacy observed under incidental learning conditions or in paradigms that make rehearsal very difficult. Enhanced recall for early list items would also result from a novelty related boost in attention and/or reduce interference from items preceding a shift in context, as would be expected either at the beginning of a list of following an event boundary.

[1] **The vast majority of subjects exhibit a primacy effect** (Healey & Kahana, 2014).

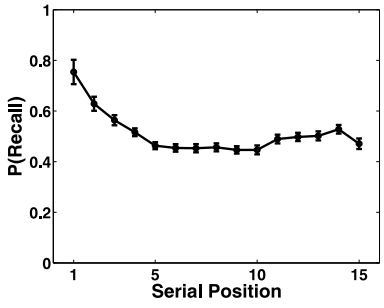

Figure 1.4: Primacy. Recall probability as a function of serial position. Error bars represent Loftus-Masson 95% confidence intervals (Loftus & Masson, 1994)

The Law of Contiguity

Scholars throughout antiquity recognized that when we experience two items, *A* and *B*, in temporal succession, subsequently thinking of *A* appears to lead us to think of *B*, and vice versa. This recognition led to a rich body of theorizing about the association of ideas (e.g., Hume, 1739), and how these associative processes may explain the secrets of human cognition. Building upon this early philosophical tradition, Ebbinghaus (1885) sought to investigate associative processes experimentally through an examination of how much more quickly he could learn lists whose sequential structure was similar to previously learned lists. This early research led to the view

that temporal contiguity was both necessary and sufficient for associative learning. Later scholars challenged these ideas by showing that without intention to learn the association among two contiguously experienced items, subjects exhibited very little evidence for associative learning (Thorndike, 1932; Hintzman, 2011). The problem, however, is that these studies may be simply demonstrating that the cognitive system can store more or less information, but it is still possible that any learning about two items will lead to some associative learning based on the contiguity principle.

Whereas early research focused on explicit tests of associative memory, such as cued recall, more recent studies have used the free recall procedure to assess the role of contiguity in memory storage and retrieval. Because the order of recall reflects the order in which items come to mind, studies of recall transitions in free recall can help to reveal the organization of memory. Kahana (1996) asked how the probability of transitioning from an item studied in serial position i to an item studied in serial position j depends on the $lag = j - i$) between the items.[2] This measure is called the *conditional-response probability as a function of lag*, or *lag-CRP*. Figure 1.5 shows lag-CRP functions obtained in three different variants of the free recall task: In immediate free recall (IFR), subjects begin recalling immediately following the final list item; in delayed free recall (DFR), subjects perform a demanding distractor task between the final list item and the recall period; in continual-distractor free recall (CDFR), subjects must perform a demanding distractor task following each and every list item.

This figure shows that in all three variants of the recall task, subjects make many more transitions among neighboring items than among items studied in more distant list positions. This is seen in the shape of the lag-CRP function, which decreases systematically as absolute lag increases, approaching an asymptotic value at moderate lags; the asymptotic value depends almost exclusively on list length, with lower asymptotic values for longer lists. The lag-CRP is also highly asymmetric, with transitions to neighbors being more likely in the forward than the backward direction. A final striking feature of the lag-CRP is the persistence of contiguity across time scales (Howard & Kahana, 1999). Although one might reasonably expect that requiring subjects to perform a demanding arithmetic task between items would sharply disrupt a subject's tendency to transition among neighboring items at retrieval, the data show otherwise; contiguity is preserved despite the disruption of the encoding process.

Figure 1.6 illustrates the generality of the contiguity effect. Figure 1.6A-C shows that the contiguity effect appears robustly for both younger and older adults, for subjects of varying intellectual ability, and for both naïve and highly practiced subjects. Figure 1.6D-F shows that the contiguity effect also predicts confusions between different study pairs in a cued recall task, in errors made during probed recall of serial lists, and in tasks that do not depend on inter-item associations at all, such as picture recognition (see caption for details). Finally, long-range contiguity appears in many real-life memory tasks, such as recalling autobiographical memories (Moreton & Ward, 2010) and remembering news events (Uitvlugt & Healey, 2019).

[2] **To compute the conditional response probabilities**, one divides the frequency of transitions to a given lag by the possible transitions to that lag, excluding transitions that are outside of the bounds of the list or transitions to already recalled items. One can also do more sophisticated corrections for autocorrelations in goodness of encoding, as discussed more fully in (Healey, Long, & Kahana, 2015).

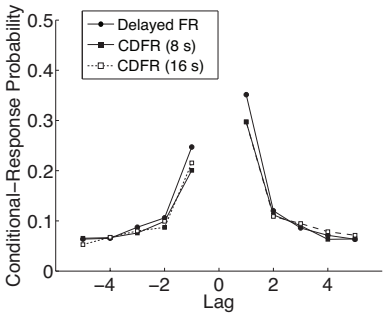

Figure 1.5: Contiguity in immediate, delayed and continual distractor free recall. The conditional-response probability as a function of lag exhibits a strong contiguity effect in both delayed and continual-distractor free recall (shown for distractors of 8 and 16 second duration). Positive values of lag correspond to forward recalls; negative values of lag correspond to backward recalls.

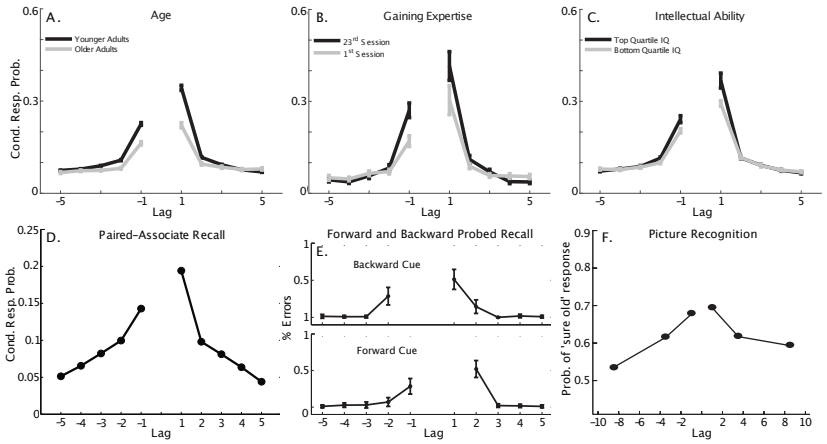

Figure 1.6: Universality of Temporal Contiguity. A. Older adults exhibit reduced temporal contiguity, indicating impaired contextual retrieval B. Massive practice increases the contiguity effect, as seen in the comparison of 1st and 23rd hour of recall practice. C. Higher-IQ subjects exhibit a stronger contiguity effect than individuals with average IQ. D. The contiguity effect appears in conditional error gradients in cued recall, where subjects tend to mistakenly recall items from pairs studied in nearby list positions. E. When probed to recall the item that either followed or preceded a cue item, subjects occasionally commit recall errors whose distribution exhibits a contiguity effect both for forward and backward probes. F. The contiguity effect also appears when subjects are asked to recognize previously seen travel photos. When successive test items come from nearby positions on the study list, subjects tendency to make high confidence "old" responses exhibits a contiguity effect when the previously tested item was also judged old with high confidence. This effect is not observed for responses made with low confidence. Healey et al (2019) provides references and descriptions of each experiment.

The Law of Similarity

Whereas the contiguity effect illustrates the temporal organization of memories, it is also well known that subjects also make use of pre-existing semantic associations among list-items (Romney, Brewer, & Batchelder, 1993; Howard & Kahana, 2002b). This can be seen in people's tendency to make recall transitions among semantically related items, even in random-word lists that lack obvious semantic associates.

As one illustration of the *semantic similarity effect*, Figure 1.7 shows how the probability of making a recall transition among two items increases with their semantic relatedness.[3] This effect is evident even at low levels of semantic similarity when lists lack any strong associates or any obvious categorical organization, providing evidence that recall transitions are driven by the relative semantic strengths among the stored items (Howard & Kahana, 2002a; Howard, Addis, Jing, & Kahana, 2007; Long & Kahana, 2017).

As a further illustration of the law of similarity we consider the everyday challenge of meeting a new person, hearing their name for the first time, and being able to subsequently recall their name when you next see them. Successful learning of name-face associations allows us to address people by their names, which is of great social significance. Sadly, many of us find it very difficult to perform these tasks without considerable effort, and even so we often have difficulty remembering a name that once came to mind with great ease. Briefly reflecting on the challenge of learning to associate names and faces reveals several potential factors that make this task particularly difficult. First, associations between names and faces are very hard to elaborate in a meaningful way. Whereas with word pairs one can often come up with a verbal mediator or an interactive image, faces are far less amenable to such elaborations, and names are rarely related to faces in any meaningful way. Second, the attributes representing the appearance of a face are likely to differ from one encounter to the next. For example, the person may have styled his or her hair differently or may be wearing sunglasses or a hat. Here we consider a third factor that is less obvious but is theoretically quite interesting and important. More than virtually any stimuli studied in the laboratory, faces are extremely similar to one another. They all have the same

[3] **Latent semantic analysis** assumes that words which are related in meaning tend to occur close together in texts. The method begins by taking a large corpus of text and counting the number of times that a given word i occurs in a given paragraph j. The resulting matrix, $L(i, j)$, has as many rows as there are words in the corpus and as many columns as there are paragraphs. A mathematical technique called *singular-value decomposition* is then used to transform the matrix in such a way as to reduce the number of columns while preserving the similarity structure among the rows. Semantic relatedness is measured by the cosine of the angle between vectors consisting of the entries in a particular pair of rows ($\cos\theta$). Completely unrelated words would have $\cos\theta \approx 0$, and strong associates would have $\cos\theta$ values between 0.4 and 1.0.

basic shape, structure, organs, etc. When you consider a set of faces representing a single gender and age range and you remove all other superficial cues (e.g., hair styling, eyeglasses, etc.), it is amazing how visually similar they are. The fact that faces are structurally so similar means that for a given name-face pair, there will be other pairs whose faces are quite similar to that of the target pair. Pantelis, van Vugt, and Kahana (2007) asked whether such similarity relations could account for why some faces are significantly easier to learn than other faces.

To assess the similarities among synthetically generated faces, Pantelis and colleagues used a technique called *multidimensional scaling* (MDS, see, Steyvers, 2004, for a review of these methods). First, they used a behavioral study to assess the pairwise similarities among faces. MDS then uses the similarity matrix derived from these ratings (or other measurements) to create an N-dimensional vector representation of each face such that the distances between pairs of faces predicts the values in the similarity matrix. With MDS we begin with a one-dimensional solution and then move to progressively higher dimensional solutions, asking at each step whether the more complex model provided a substantial improvement in the models' ability to fit the similarity matrix. Pantelis et al (2007) found that a 4-dimensional solution (each face described by a set of 4 numbers) provided a very good fit to the similarity matrix. The output of the model was a a vector representation of each face in a four-dimensional space, where each dimension roughly corresponded to one of the major attributes of the faces (e.g., the shape of the hairline; the width of the nose). Defining each face as a vector allowed them to calculate the similarities among any pair of faces.

Pantelis et al. hypothesized that people would have greater difficulty associating names with faces that had many "neighbors" in the four-dimensional face space (where neighbors are defined as the number of other faces that lie within a small radius around the target face). To test this hypothesis they asked participants to study eight faces that were paired with common American male names. At test, participants were presented with each of the eight studied faces, one at a time (see Figure 1.8A). As each face appeared, participants attempted to recall the name that was previously paired with the face. This study-test procedure was repeated 10 times so that participants could learn the names of all of the faces.

When cued with a face at test, participants' accuracy at recalling the correct name decreased as the the number of also-studied neighboring faces increased. RT data showed the inverse effect (see Figure 1.8B). Also, when participants made intralist intrusions, recalling a name paired with a different face on the target list, these errors tended to be names associated with faces that were similar to the target face (see Figure 1.8C).

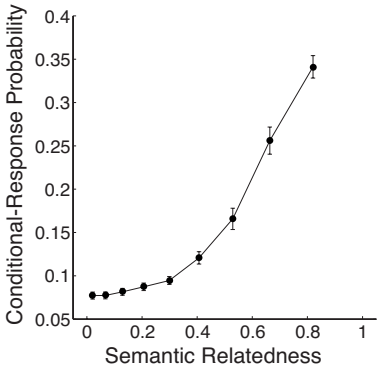

Figure 1.7: Semantic similarity effect. Subjects' recall transitions favor semantically similar items. Latent semantic analysis provided a measure of relatedness between 0 and 1 (Landauer and Dumais, 1997). Data from Kahana and Miller (2013).

The Law of Repetition

Notwithstanding the well-known adage that "Practice makes perfect," the earliest scholars to study memory in the laboratory soon discovered that some forms of practice are better than others. Both Ebbinghaus (1885) and Müller and Schumann (1894) reported that repetition produced more rapid learning when the repeated trials were spaced apart as compared with when they were massed together.

Figure 1.9 illustrates the spacing effect in two classic recall paradigms:

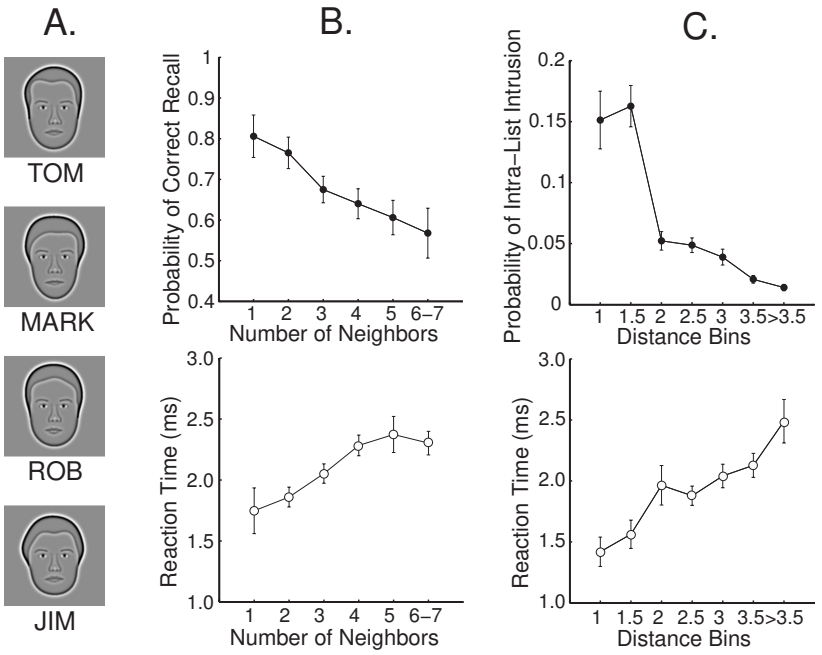

Figure 1.8: Similarity and memory for name-face associations. A. Examples of name-face pairs presented at study. **B.** Neighborhood Effect. The upper panel shows the probability of recalling the correct name when cued with a face at test, as a function of how many neighbors that face had within the study list. The lower panel shows response times for correctly recalling names of faces as a function of the number of neighbors. **C.** Each possible intrusion name corresponded with a study face, for which the distance from the cue face in four-dimensional face space was calculated. The upper panel shows the probability of making an intralist intrusion of a particular distance. The lower panel shows the response times for intralist intrusions of various distances from the cue face.

cued recall and free recall. Madigan (1969) had subjects study lists comprising a mix of once-presented (1P) and twice-presented (2P) items. The 2P items were either presented successively (spacing = 0), or with varying numbers of items separating the repetitions (2, 4, 8, 20, or 40). Immediately following presentation, subjects were asked to freely recall the words in any order. Figure 1.9A shows that increasing the separation of the 2P items improved their recall. Figure 1.9B illustrates a similar advantage for spaced practice in a paired-associate learning task. In this study, Glenberg (1976) had subjects study randomly-paired words with some pairs repeated at varying lags. He found an advantage for spaced practice that extended out to 20 or more intervening pairs.

Spacing effects have also been shown in real-world situations such as learning to type (Baddeley & Longman, 1978), learning foreign languages (Bahrick & Phelps, 1987), and remembering television commercials (Singh, Mishra, Bendapudi, & Linville, 1994). In the case of commercials, companies pay handsomely to present a brief message to a large audience in the hopes that the viewers will remember that message. Thus, it is not only important to produce a compelling message, it is also important to present the message in a way that maximizes viewers' memory for this incidentally encoded material. In a study by Singh et al. (1994), 413 people, including both young and older adults, viewed a television news program during which several commercials were presented. Some commercials were presented once, and some were repeated. The repeated commercials were separated by either one or four non-repeated commercials (spaced short and spaced long conditions). The researchers told participants that they were about to see a late-night news show taken from a network-affiliate station outside their own viewing

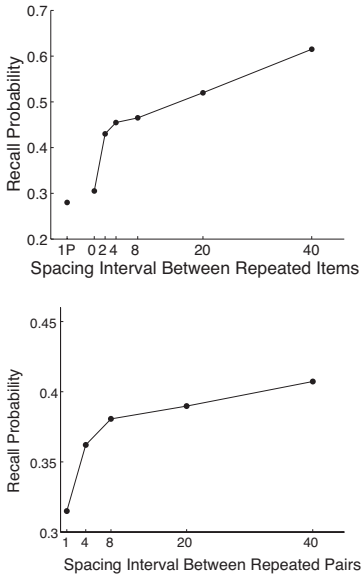

Figure 1.9: The beneficial effects of spaced practice. A. Probability of freely recalling words increases as a function of the spacing (lag) between their repeated presentations. 1P=once presented items (Madigan, 1969). **B.** Cued recall probability also increases as a function of the spacing of repetitions (Glenberg, 1976).

area and that they were interested in knowing the participants' opinions about the news show. The following day, participants were brought back to the lab and tested on their memory for the commercials. They were given the name of a product category and asked to recall the brand name and claims made by the respective commercials. Participants' recall was 18% higher when the repeated commercials were spaced by four non-repeated commercials than when they were spaced by just one commercial. This advantage of spaced learning is one of the most robust and practical findings in the human memory literature.

Processes

The foregoing analysis highlighted five very general principles of memory: the so-called laws of recency, primacy, contiguity, similarity and repetition. Any successful theory of the processes underlying memory encoding and retrieval must somehow embody these five principles.

Systems vs. Processes

Two major cognitive approaches predominate in the study of human memory. The first of these approaches, termed the *memory systems approach*, assumes that memory is not a singular entity, but rather is an umbrella term for a web of brain systems that each support different kinds of memory function. The goal of this approach is to determine what the different memory systems are and to identify the brain regions supporting each of these systems. Much as a biologist wants to classify the different types of organs that make up the body, the different types of cells that make up an organ, and the different types of processes that regulate cell function, memory researchers seek to identify the major subtypes of memory, the brain systems that give rise to them, and how these systems interact with one another. The second major cognitive approach involves the construction of computational models that describe the processes underlying memory function. Following this approach, researchers attempt to understand memory by specifying mathematical equations that characterize the encoding and retrieval processes underlying memory function. Having specified these equations, a computer is then used to solve the equations and thus generate predictions for what should happen in a memory experiment.[4] With these predictions in hand, scientists can ascertain where the model succeeds and where it fails to match the experimental data. These successes and failures give scientists ideas as to which of the model's assumptions are flawed, and how to fix them. Sometimes, the data needed to test a crucial aspect of a model do not exist. In these cases, an experiment must be designed to test the model. This interaction between data and models helps scientists create models of human memory that can account for a broader array of known data and that can also make better predictions about new memory phenomena that are yet to be observed.

During the past two decades a transformative development has taken place in the study of human memory. For the first century of memory research, models and theories of memory have been aimed at explaining experimental data on overt behavior, such as accuracy and RT in recall or recognition-memory tasks. Starting in the 1990s, a major emphasis in mem-

[4] In rare cases, one can solve the equations of a memory model analytically, as you would an algebra or calculus problem. However, most memory models are too complex to solve without the aid of a computer.

ory research has been to understand the biological mechanisms responsible for the encoding, storage, and retrieval of memories. In the next chapter we will turn to the biological foundations of human memory, which will set the stage for the analysis of electrophysiology through the rest of the book.

Representation and Search

A common assumption in many theories of memory is that our brains parse the incoming stream of experience into specific units which can be stored, retrieved, and combined into higher order units. This idea leads to the concept of *representation*—the idea that a given unit of memory can be represented by a mathematical object. Whereas early scholars conceived of each memory as an indivisible node in an associative network, subsequent scholars raised the possibility that each memory is actually a bundle of smaller units called elements, features, or attributes. We will refer to this as the *distributed representation hypothesis* (as distinct from a localist representation where a single unit represents a single complex memory). The distributed representation hypothesis can be traced back at least to the writings of the great scientist/inventor Robert Hooke (1969), but these ideas also emerged in the work of philosopher/biologist Richard Semon (1923) and the writings of the early learning theorist Edwin Guthrie (1935).[5] The distributed representation recording a particular experience in memory is sometimes called the memory *trace* or, after Semon (1923), the *engram*. During the mid-20th century, mathematical learning theorists developed formal models of memory based on the ideas of distributed representations (e.g., Estes, 1955; Bower, 1972), and these models provided the foundation for important subsequent work on neural network theories of memory (McNaughton & Morris, 1987; Rumelhart, McClelland, & the PDP Research Group, 1986)

[5] Hintzman (2003) and Schacter (2001) provide fascinating discussions of the work of Hooke and Semon, respectively (see, also, Gomulicki [1953]).

For simple geometric forms (e.g., two-dimensional rectangles), the dimensions of items can be specified explicitly, and these physical dimensions often turn out to be the same as the psychological dimensions along which the items vary (e.g., Nosofsky, 1992; Kahana & Bennett, 1994). In the case of memorizing words or complex pictures, however, the theories assume that the items vary along a great number of physical and psychological dimensions whose identities may be difficult, or impossible, to fully identify. Latent Semantic Analysis, described above, provides one way of conceptualizing vector representations of words.

Another way to conceptualize the vectors representing complex stimuli is to think of them as characterizing the pattern of brain activity evoked by processing a given stimulus. Ultimately, all stimuli must be represented by the electrical activity of neurons in the brain. Some of these neurons may be very active, exhibiting a high firing rate, whereas others may be quiet. The firing rate of each neuron can be thought of as representing the value along the attribute coded by that particular cell (in reality a large number of neurons are likely involved in the coding of any attribute, not a single cell). Together, these cells can represent many kinds of information, including perceptual, contextual, and semantic aspects of an experience. Much as an image on a television screen is a pattern of brightness values distributed across the display, a memory may be thought of as a pattern of neural activation values distributed over a large array of neurons.

Assuming that some desired information is stored in memory, how does

one find it? This *search problem* is particularly challenging in memory tasks where the to-be-remembered items are already well learned, such as the words in one's own language. In such tasks, one must still remember which words appeared on the list. To understand how people solve this search problem, one must not only specify the representation of the items and the processes governing the encoding of these representations in memory, but one must further specify how items are retrieved. Hypothesized retrieval mechanisms, which are central to all contemporary theories of memory, typically rely on the concept of association and the related concept of *cue-dependent retrieval*. In these theories of memory, different stored representations can evoke one another via associative connections formed both during the study of a target list and during one's prior experience with the studied items. These associations can be used to retrieve specific memories despite their being blended with many other non-targeted memories.

Multiple Traces

With distributed memories, each item is represented by a vector of values, one for each dimension or attribute, and the set of all items in memory can be thought of as an array where each row represents one dimension and each column represents a different item. Such an array of values is a *matrix*.

Here we introduce a simple yet powerful model for the learning process: a model in which each studied item lays down a new trace in a large and ever-growing array representing all of the traces stored in memory. This idea, which is referred to as the *multiple trace hypothesis*, implies that each encoded presentation of an item leaves its own memory trace (Hintzman, 1976; Moscovitch, Nadel, Winocur, Gilboa, & Rosenbaum, 2006). By allowing each studied item to lay down a unique trace, and by further assuming that each trace can consist of many attributes, we can easily accommodate the important idea that a given item (e.g., the word *cat*) will lay down somewhat different traces when studied on different occasions. Indeed, it seems strange to think that the encoding of a given word will be precisely the same at any two occasions. A much more natural assumption is that the stored attribute values vary based on the context in which a word occurred, such as the words that preceded it or the thoughts that it evoked.

The multiple trace hypothesis implies that the number of traces can increase without bound. Although the limits of information storage in the human brain are not currently known, it seems implausible for the brain to have an infinite storage capacity. The presence of an upper bound need not pose a problem for the multiple-trace hypothesis so long as traces can be lost/erased, similar traces can merge together, or the upper bound is large relative to the scale of human experience. We next consider how we can address the process of memory search in a model in which memories are stored in a matrix representation.

Summed Similarity

When a person encodes a test item, we assume that it is converted into a vector representation, which can then be compared with all the vectors stored in the memory matrix. One way to recognize an item is to search the memory matrix, serially comparing the probe vector with each of the stored

The matrix encompassing item-vectors $\mathbf{m}_1$, $\mathbf{m}_2$, and $\mathbf{m}_3$ is:

$$M = \begin{pmatrix} m_1(1) & m_2(1) & m_3(1) \\ m_1(2) & m_2(2) & m_3(2) \\ \vdots & \vdots & \vdots \\ m_1(N) & m_2(N) & m_3(N) \end{pmatrix}$$

vectors. Although such a search process may be plausible for a very short list of items, it has difficulty accounting for the very rapid speed with which people can recognize test items drawn from long study lists.

As an alternative, one can search the memory matrix in *parallel*, comparing the probe item with each of the stored vectors at the same time. In carrying out such a parallel search, we would say *yes* if we found a perfect match between the test item and one of the stored memory vectors. However, if we required a perfect match to say *yes*, we may never say *yes* because a given item is likely to be encoded in a slightly different manner on any two occasions. Thus, when an item is encoded at study and at test, the representations will be very similar, but not identical. To circumvent this problem, we could accept partial matches so long as they exceeded some threshold of similarity. Alternatively, we could calculate the similarity for each comparison and sum these similarity values to determine the *global match* between the test probe and the contents of memory. Models that adopt this approach are called either *summed-similarity* or *global-matching* models.

Several versions of the summed-similarity approach have been proposed in the literature, and these models have proven to be quite successful in accounting for a wide range of recognition-memory data. Before we consider these models in greater detail, however, we need to address the issue of how to focus the search process on those items learned within a given context. In most memory experiments, this context is the most recently presented list.

Temporal Context

Laboratory studies have shown that changes in *situational context* between study and test can reduce performance on a range of memory tasks. These findings led to the view that contextual change is one of the major factors underlying forgetting (Carr, 1931; Hollingsworth, 1928; McGeoch, 1932; Robinson, 1932). Models of memory can account for this finding by allowing a subset of the attributes that represent an item to represent the situational context in which an item was learned, including its time and place of encoding. When context is not specifically manipulated in an experiment, by changing the location or other background attributes between study and test, we will use the term *temporal context* to refer to the contextual attributes that changed. Of course these attributes need not represent time per se; rather they represent any internal representation that accompanies our experience that varies with time, including background thoughts, affective states (i.e., are you happy or sad), and physiological variables (e.g., hunger, tiredness, anxiety).

By allowing for a dynamic representation of temporal context, items within a given list will have more overlap in their contextual attributes than items studied on different lists, or indeed items that were not part of an experiment (Bower, 1972; Crowder, 1976). If the contextual change between lists is sufficiently great, and if the context at time of test is similar to the context encoded at the time of study, then recognition-memory judgments of the type described in the section on *Memory Paradigms* should largely reflect the presence or absence of the probe (test) item within the most recent (target) list, rather than the presence or absence of the probe item on earlier lists. This enables a multi-trace summed-attribute similarity model to account for many of the major findings concerning not only recognition memory tasks,

but a number of related tasks involving memory judgements.

Summed-similarity Computations

Within the framework of attribute theory, two memories are identical if they share the same values along each attribute. Intuitively, the similarity of two vectors should decrease as the distance between them increases.[6] When two vectors are identical (i.e., the distance between them is zero), their similarity should be set to some maximal value. As the distance between the vectors increases toward ∞, their similarity should approach zero. The exponential decay function,

$$\text{similarity} = e^{-\tau \text{ distance}}$$

has exactly this property: it is equal to 1.0 when distance $= 0$ and it approaches zero as distance $\to \infty$. The variable τ (Tau) determines how quickly similarity decays with distance[7]. Armed with measures of similarity, we can sum the similarity between a test item and each of the stored vectors in our memory matrix. Let us assume that a person says "yes, I remember that item" if the summed similarity exceeds a threshold value.

To formalize these arguments, suppose that $\mathbf{m}_i$ represents the ith item of an L-item list. Following study of the list, the matrix M would represent the list in memory.

$$M = \begin{pmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \mathbf{m}_3 & \dots & \mathbf{m}_L \end{pmatrix}$$

Now let $\mathbf{g}$ represent a test item, either a target (i.e., $\mathbf{g} = \mathbf{m}_i$ for some value of i) or a lure. The summed similarity between the test probe and the items stored in memory can be written as:

$$\sum_{i=1}^{L} \text{similarity}(\mathbf{g}, \mathbf{m}_i), \tag{1.1}$$

where the similarity between $\mathbf{g}$ and $\mathbf{m}_i$ is defined as:

$$e^{-\tau ||\mathbf{g} - \mathbf{m}_i||} = e^{-\tau \sqrt{\sum_{j=1}^{N} (g(j) - m_i(j))^2}}. \tag{1.2}$$

The summed-similarity model dictates that a person will respond *yes* when the summed similarity between the test item and each of the stored items in memory (Equation 1.1) exceeds a threshold value C; otherwise the person responds *no*. The threshold value can be set to simultaneously maximize hits (correct *yes* responses to studied items) and minimize false alarms (incorrect *yes* responses to non-studied items).

Modeling Context Dynamics

Within the attribute similarity framework, we can account for the *Law of Recency* by assuming that the contextual attributes change slowly as each new item is presented (e.g., Estes, 1950, 1959). How might such a contextual-drift process work? Let us first denote the set of attributes representing context as the vector $\mathbf{t} = (t(1), t(2), \dots, t(N_{\text{context}}))$, where N_{context} is the number of attributes representing context. Now suppose that each time a new item is presented or remembered, we change the context vector by adding a small variable amount to each element. We can thus write down a simple Gaussian random-walk model for the context vector as:

$$\mathbf{t}_i = \mathbf{t}_{i-1} + \epsilon \tag{1.3}$$

[6] The distance between two vectors, $\mathbf{m}_1$ and $\mathbf{m}_2$, is the length of the difference vector, $\mathbf{m}_1 - \mathbf{m}_2$, which is $\sqrt{\sum_{i=1}^{N} (m_1(i) - m_2(i))^2}$, where N indicates the number of attributes or dimensions.

[7] A related measure of similarity is the cosine of the angle between two vectors, $\cos\theta$. If the two vectors are identical, the angle will be zero and $\cos\theta = 1$. If the two vectors are perpendicular to one another, the angle will be 90° and $\cos\theta = 0$. If we were to select random values for each of the attributes, we would find that $\cos\theta = 0$ in expectation.

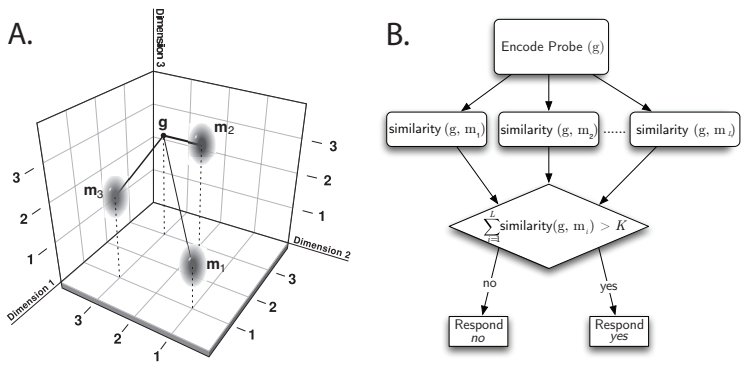

Figure 1.10: Illustration of the Summed Similarity Model. **A.** Schematic showing the memorial representations of three list items in a three-dimensional attribute space. Each item ($\mathbf{m}_1$, $\mathbf{m}_2$, and $\mathbf{m}_3$) is depicted by a shaded ellipse representing the noise associated with the coding of the item. Also shown is the relatively noiseless representation of a nonstudied test probe (**g**). **B.** Diagram of the information-processing stages in the summed similarity model. First, the test probe is encoded. Then, its similarity to the memorial representations of each list item is computed. If the summed similarity exceeds a threshold, C, the model responds *yes*.

where ϵ is a random vector whose elements are each drawn from a Normal, or Guassian, distribution, and where i is an index variable that counts each item presentation. Assuming that context changes gradually over the course of an experiment, the amount of change in context between the study of an item and its later test will increase with the number of items intervening between study and test. This is how context can be used to explain the recency effect. Recent targets will have higher summed similarity than remote targets (See Figure 1.11).

Summed-Similarity Summary

The summed-similarity framework provides a comprehensive account of many findings in the recognition memory literature. It embodies the principles of similarity, recency, and repetition mentioned above. The basic form of the model discussed here does not account for contiguity-based associations or the principle of primacy. More importantly, however, summed-similarity derives a single number (a scalar) to represent the output of the memory system. Such an approach fails to account for our ability to recall specific items or the cue-dependent nature of the recall process. The essential missing element is a process model of association. We next consider associative network models that account for both association and recall dynamics.

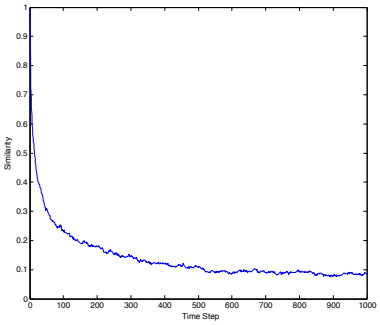

Figure 1.11: Contextual similarity decreases with each successively processed item. Similarity is defined as the $\cos\theta$ (**t**, **t**+τ) where τ is the number of time steps.

Neural Networks

A major advance in our understanding of associative memory took place in the 1970s when computer scientists, neuroscientists, and psychologists discovered mathematical methods for describing how associations could be represented in memory, and retrieved in a cue-dependent manner. These models of associative memory were inspired by the circuitry of the brain itself, where connections between neurons appear to play a crucial role in associative learning and recall. This class of models developed a strong following among scientists in diverse fields, and the research that grew out of this work is known as *connectionism* or *neural networks*.

Neural-network models assume that each neuron (or node) is connected to many other neurons. Together, these neurons (nodes) form a highly interconnected network. In our model system, each neuron can be characterized

by its *activation* value, and each connection is characterized by its strength value, or *weight*. At a given time, the vector of activations for the neurons in the network is called the *state vector* of the network. The state vector, which can represent the attributes of a memory, is not static. It changes from moment to moment as our mind wanders or as new information is being experienced. So, although the state vector can represent a previously stored memory vector, it does not store any memories. Rather, memories are stored in the connections between neurons, and these connections enable the network to recover previously stored memories. The connection strengths are also not static. They change during learning to allow the state vector to recall previously learned memories. An important property of neural-network models is that they can store a large number of distinct memories in the pattern of connections among the nodes of the network.

Network Dynamics

Each node, or neuron, in a neural network gets input from other nodes and sends output to other nodes. A scalar value, termed activation, describes the output of each node. The activation of node i, denoted $a(i)$, is related to the input that it receives from other nodes in the network. The input that node i receives from node j is the product of $a(j)$ and the weight of the connections between i and j, denoted $w(i,j)$. Putting this together, we can write:

$$a(i) = g\left(\sum_{j=1}^{N} w(i,j)a(j)\right) \tag{1.4}$$

where the function $g(\text{input})$ transforms the unit's input into its activation value. For now we will assume that the activation of a unit equals its input (i.e., $g(\text{input}) = \text{input}$). This type of simple neural network is called a *Linear Associator*.[8] Equation 1.4, proposed by McCullough and Pitts (1943), is called the *dynamical rule* of the network. This rule provides a simplified description of the behavior of our artificial neurons, whereby activations map onto the neurons' firing rates and weights map onto the strength of the synaptic connections between the neurons (see Figure 1.12).

[8] In modeling the behavior of a neural network, it is useful to define a neuron's activation as being either above or below its average activation level. This is done by defining an activation of zero as the average activation of a neuron: positive activation values would denote above-average activity and negative activation values would denote below-average activity.

Hebbian Learning

Hebb (1949) proposed that learning results from changes in the connection strengths, or weights, between neurons. Specifically, he hypothesized that the weight of the connection between neurons i and j (denoted $w(i,j)$) is increased by the product of their activity at each time-step t. Mathematically, we can write the Hebb learning rule as:

$$w(i,j)_t = w(i,j)_{t-1} + a(i)_t a(j)_t, \tag{1.5}$$

where $a(i)_t$ denotes the activation of neuron i at time t. Although it is convenient to assume that all of the weights are initially set to zero (when $t = 0$), one could also assume that the weights are initially set to some random values.

According to the Hebb rule, if neurons i and j both have above-average (i.e., positive) activations, their connection is strengthened. Similarly, if they both have below-average (i.e., negative) activations, their connection is also

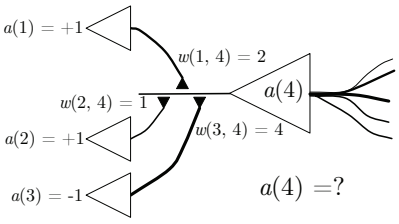

Figure 1.12: McCullough-Pitts model neuron. Neurons (open triangles) receive input from the left and send output to the right (this arrangement is arbitrary). The activation of Neuron 4, $a(4)$, is determined by the weighted sum of the inputs coming from Neurons 1–3 (in this example $g(-1) = -1$).

strengthened. But, if one neuron's activation is positive while the others is negative, their connection is weakened.

Modeling Recall

Consider two sets of neurons: one set represents the attributes of item A and the second set represents the attributes of item B. Let $a(i)$ denote the output of the ith unit representing A and let $b(j)$ denote the output of the jth unit representing B. (We again assume that items are vectors of attributes, but we now assume that each attribute value is coded by the firing rate of one of our simplified neurons). Suppose, further, that each neuron representing attributes of A is connected to each neuron representing attributes of B. We can now write the weight matrix as:[9]

$$w(i,j)_t = w(i,j)_{t-1} + a(i)_t b(j)_t. \tag{1.6}$$

[9] In matrix notation, the learning rule in Equation 1.6 is given by:

$$W_t = W_{t-1} + \mathbf{b}\mathbf{a}^\top$$

where **a** and **b** are column vectors.

To see how associative (cued) recall works, set the activations of the neurons representing A to their appropriate values and allow Equation 1.4 to determine the activations of the neurons representing item B. The output (activation) of the jth node of B is given by:

$$\tilde{b}(j) = \sum_{i=1}^{N} w(i,j)a(i) = \sum_{i=1}^{N} a(i)b(j)a(i) = b(j)\sum_{i=1}^{N} a(i)^2, \tag{1.7}$$

where $\tilde{b}(j)$ is the new value of $b(j)$ after it has been updated by applying the dynamical rule. If we assume that each item vector, **a** and **b**, is of length one, then it is easy to show that $\tilde{\mathbf{b}} = \mathbf{b}$ because $\sqrt{\sum_{i=1}^{N} a(i)^2} = 1$.

Now if instead of just learning a single pair of items, we stored many pairs of items within the same neural network. It is easy to show that the network can still recall the correct item so long as all of the items are orthogonal vectors. Sparse, random vectors of high dimensionality will be very nearly orthogonal, so even a linear associator can accurately recover a good approximation of the target memory. If, however, you want to store and retrieve associations among similar items, as in the name-face study described above, you would need to move to a non-linear activation function to enable the network to iteratively converge to the correct target memory. This process if often called *deblurring*, and there are many types of networks that can achieve a high storage capacity for associative memories while still accurately recovering the target association.

Neural Network Summary

Inspired by the computational properties of actual neurons, neural network models allow memories to be evoked directly by a cue item without requiring a conventional search process. As such, the stored memories are referred to as being "content addressable." Neural networks store memories, defined as patterns of neural activity, in the strengths or weights of the connections between neurons. One of the most important and attractive features of neural networks is that they can store a multitude of memories in a single set of connections. As each new memory is experienced, the weights connecting the network's neurons are updated according to a learning rule. We focused on one such learning rule, known as Hebbian learning, whereby the weight

between two neurons increases or decreases as a function of the product of the neurons' activations.

Rather than directly comparing a probe item with each of the stored memories, the retrieval (or recall) process in a neural network occurs as a natural process of the networks dynamics. When the network is not in a learning mode, each neuron's activity depends on the activity of the other neurons in the network, and the connection weights to those neurons. The simple dynamical rule illustrated in Figure 1.12 on page 27 and defined by Equation 1.4 on page 27 allows the network to "recall" previously learned patterns.

As with summed-attribute-similarity models, one can incorporate contextual features into neural networks to distinguish between identical items encoded on distinct occasions or in different contexts. Assuming that context changes slowly over time, the network embodies the Law of Recency. However, without adding some additional assumptions or mechanisms, neural network models can not account for the beneficial effects of spacing on subsequent cued recall.

Organization of the Book

The present chapter introduced some of the major paradigms, principles and processes investigated by students of memory. List recognition and recall tasks, involving their distinct study and test phases, have fueled much of the theorizing on human memory and these tasks (and their variants) continue to occupy center stage. In later chapters we will discuss a number of other memory paradigms including spatial learning and spatial memory, category learning, and probability learning. Citing data from recognition and recall tasks we documented five major laws of memory: recency, primacy, contiguity, similarity, and repetition. Finally, we surveyed some of the major process models of memory, including summed exemplar similarity models and connectionist network models. Absent from chapter 1 was any discussion of electrophysiology, which is the main focus of this book. The foundational background to the electrophysiological analysis of memory appears in the two remaining chapters of Part 1: Chapter 2 surveys Neurons, Fields and Networks, focusing on key ideas from neurobiology; Chapter 3 introduces the field of human electrophysiology, surveying the major methods and presenting some key findings. Chapter 3 also introduces two large public datasets that we analyze in later chapters.

Human electrophysiology is a big data affair. The size of a single dataset can often exceed one terrabyte. Parts 2 and 3 of the book provide a hands-on tutorial to the key data science methods used in the analysis of human electrophysiological data. Part 2 begins by introducing the event related potential technique, which allows researchers to identify electrophysiological signals that are time-locked to the presentation of a stimulus, or the execution of a motor response. Next, we introduce the reader to spectral analysis methods that have been used to uncover the role of oscillatory or rhythmic activity in human learning and memory. These techniques allow researchers to transform a complex time varying signal from the time domain to the frequency domain and in doing so, relate signals such as an evoked potential to the alignment of oscillatory activity. We also discuss how spectral methods can be used to analyze non-oscillatory activity, which may nonetheless exhibit

correlations with behavior. The final chapter in Part 2 discusses single cell recordings and their use in the study of human memory. Part 3 introduces multivariate approaches to the analysis of electrophysiological data, focusing on regression and classification models that form the foundation of modern machine learning (ML). We show how these ML techniques have allowed researchers to decode the content of human thought, bringing extrasensory perception from fiction to reality. In the final chapter of Part 3, we review methods for studying the functional connectivity underlying memory and cognition. Here we discuss both spectral and statistical modeling methods used to assess functional connectivity. In part 4 of this volume we survey several major applications of electrophysiological methods to the study of human memory and cognition. This section includes chapters on verbal memory, spatial memory, recognition and working memory, and brain stimulation.

2
Human Electrophysiology

Historical Background

Technological innovations that improve the measurement and quantitative analysis of data drives scientific progress. Indeed, we can identify four developmental phases in the evolution of human electrophysiology that arose from specific technological advances. The first phase followed Hans Berger's famous discovery of the human electroencephalogram (EEG) in 1929. By applying electrodes to the scalp Berger was able to identify 10 Hz rhythmic activity overlying the human occipital cortex—the so-called alpha rhythm. Following the discovery of this so called alpha rhythm several other rhythms were identified, such as the 1-2 Hz (delta) rhythm associated with sleep and the 10-20 Hz rhythm overlying motor regions of the brain that increase during movement. By the late 1930s the use of the EEG had become widespread in clinical neurology and psychiatry, and many researchers began to explore the EEG correlates of both disease and normal cognitive function. For a fascinating and erudite analysis of research on brain rhythms, including many historical side notes, we refer to the reader to Buzsaki (2004).

The emergence of the next major phase of electrophysiological research coincided with the widespread adoption of mainframe computing machines by major academic institutions. These computers made it feasible for researchers to analyze large data electrophysiological datasets using statistical methods. (Psychophysiologists were early adopters of computing machines, such as the IBM 360 and the Digital Equipment Corporation PDP series of computers). Although scientists applied many data-analytic methods to EEG data, the evoked-potential method came to dominate research in this period, leading to the identification of time-locked modulations of the EEG, known as event-related potentials, or ERPs. One prominent example of these was the P300, a positive going ERP component appearing at around 300 milliseconds following an auditory stimulus. Chapter 4 describes the ERP method in detail, including discussion of the P300 and other components. The analysis of event-related potentials dominated the physiological study of human memory and cognition for more than two decades.

Although the second phase of research involved moving psychophysiological research out of the neurological clinic and into the psychological laboratory, the third phase involved a return to the clinic where researchers could gain access to invasive electrical recordings taken from neurosurgical patients being treated for drug-resistant epilepsy. Although researchers

Figure 2.1: The discoverer of the human electroencephalogram, Hans Berger (1873 – 1941) was a professor of neurology at Jena, in Germany.

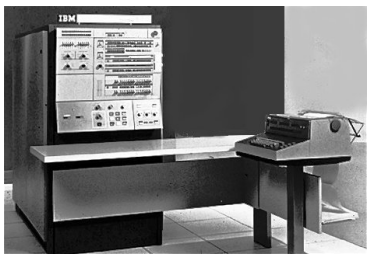

Figure 2.2: Photograph of the IBM360 mainframe computer.

had been studying neurosurgical patients with epilepsy throughout the earlier periods, research in the clinic was not for the faint of heart. Until the late 1990s, EEG systems being used in clinics relied on archaic analogue recording systems and lacked any convenient means of synchronizing between the recordings taken and the computers that controlled the memory experiments. Developers of these proprietary systems wanted to create a bullet-proof system for clinical use and did not aim to make their data easily readable or accessible to scientists. Few psychophysiologists were able to conduct research under these circumstances and the vast majority of data were therefore generated using scalp-EEG recording systems optimized for basic research and employed in studies of healthy individuals.

The third phase of human electrophysiological research involved a return to the roots of the field, with researchers setting up controlled experiments in the epilepsy clinics to gain access to precious electrical recordings from deep brain structures, such as the hippocampus. These invasive brain recordings led to a re-evaluation of the rhythmic components of EEG signals. These developments served the critical role of bridging human and animal work on the electrophysiology of memory, leading to greater dialogue between neurobiologists studying memory in rodent models and cognitive neuroscientists studying memory in humans.

This third phase of research would not have been possible were it not for important technological developments in personal computing. Although the PC revolution began in the 1980s, it took another decade or so before these desktop machines replaced mainframe systems for most scientific computing needs. By the mid 1990s, advances in computer speed and the emergence of scientific computing platforms, such as MATLAB, allowed researchers to perform sophisticated quantitative analyses of EEG data with relative ease[1]. Researchers could use these new software tools instead of relying on proprietary systems mostly tailored to conducting ERP analyses. This in turn led to a dramatic increase in the sophistication of analytic methods, and in particular to a surge of interest in spectral analysis methods that we will discuss in Chapter 5. These spectral methods allowed researchers to return to the analysis of rhythmic brain activity with far greater rigor than was possible in the first half of the 20th century.

At the dawn of the 21st century there were dozens of research groups collecting and analyzing direct brain recordings in epilepsy centers around the world. These recordings included not only field potentials, but also signals from individual neurons, and researchers became very busy examining how these brain signals correlated with a host of perceptual and cognitive variables. This period also saw a number of striking demonstrations of conserved neural signals between animals and humans, particularly in the realm of spatial exploration, as discussed in Chapter 14. Researchers studying non-invasive scalp EEG also embraced spectral methods and the tools needed to perform these analyses were rapidly developed and freely disseminated.

Perhaps it is too early to identify a fourth phase in the evolution of the study of human electrophysiology. Nonetheless, the signs of significant change are upon us. Human brain recordings produce immense datasets, and the ability to make statistical inference on these data sets requires that we build multivariate models while avoiding overfitting our data. Although statistical methods for solving these problems have been known for decades, it is only in recent years that we have had the computing power to deploy

[1] EEG data form a series of voltage time series measured from different locations in the human brain or on the scalp. Spectral analysis methods may be used to measure both the presence of oscillations within a single time series as well as the phase locking of oscillations across two different voltage time series. Spectral analysis of EEG data reveals two important features of the time series of EEG data. First, EEG data, like most natural processes, has more energy at low than at high frequencies. This is a consequence of the autocorrelation within the time series; that is, the voltage at time t depends on the voltage at time $t-1$. The second feature of the time series is the presence of increased amplitude (or power) at specific frequencies. The presence of peaks at specific frequencies implies that, in addition to its generally slow-changing voltage pattern, the EEG time series can also contain oscillations at specific frequencies.

these "machine learning" methods on large datasets. A principal goal of these methods is to ask whether the brain data we have recorded has information about some behavior or cognitive state, such that we can use the neural signals to make predictions in an independent sample. This approach has significant practical utility in that it can be used to directly decode behavior or cognitive states from brain activity. The chapters in Part 3 of this book provide a hands-on tutorial for using these methods, and the chapters in Part 4 illustrate the application of these techniques in several cognitive domains.

As has been widely discussed (Millett, 2001), Hans Berger sought to find a physiological basis for mental telepathy. He pursued this goal for decades before he finally convinced his colleagues that he had found a reliable electrical signal emanating from the human brain. Today we are using multivariate statistical models to decode the variations in rhythmic activity throughout the brain to enable patients with severe motor impairments to communicate with loved one's using brain signals, a feat not too far from the kind of mental telepathy that Hans Berger sought to understand.

Anatomy of an Experiment

Consider a list memory experiment in which subjects study items for a subsequent recognition or recall task. A researcher would normally program a computer to create the lists according to a set of pre-defined rules, present the lists either visually or auditorally to the subject, and record the responses. In these experiments subjects either give multiple-choice responses (e.g., Yes/No) or they provide vocal or typed responses to recall a word presented on a studied list. The computer would also create a log file to insure that all experimental events are recorded and time-stamped with an accurate algorithm to insure that the time stamps recorded to file closely match the true times at which the events occurred during the experiment. Using standard programming languages (e.g., Python, C++) coupled with some specialized experiment programming toolboxes, it is relatively straightforward to program a personal computer to control almost any experiment.

When you record physiological data, such as EEG signals from the scalp or direct recordings of field potentials from inside the brain (including single neuron responses), there is a separate device, also controlled by a (different) computer that manages the logging of the physiological signals. Any physiological study of memory requires an accurate method for aligning the physiological and behavioral data, so that one can create a final data structure with a single set of time stamps. [2]

When recording electrophysiological data there are two essential decisions that must be made at the outset: 1) The sampling rate at which the system will record the signals, and 2) the way that the electrical signals will be referenced. We briefly discuss each of these considerations below.

Sampling Rate. Higher sampling rates are akin to having more decimal places when you do math, and you may think that the higher the sampling rate the better. However, if the signals of interest are measured at the time scale of milliseconds there is no need to measure signals at nanosecond resolution. For scalp EEG signals, it is sufficient to record data at 1kHz (i.e., one sample per millisecond). However, for intracranial recordings with

[2] Whereas EEG systems tailored for recording evoked potentials sometimes default to only record during specific behavioral epochs, it is much better to record continuous physiological data throughout the entire recording session. This can substantially improve your ability to filter the data, either to remove noise sources or identify slow rhythmic components in the data.

microwires, resolving the firing of individual neurons requires sampling rates of 10 kHz or higher.

Reference scheme. As described in Chapter 2, neural activity supporting cognition produces the movement of charged particles (ions) in the brain. As such, the distribution of charge particles, and their associated *electrical fields*, result in changes in the *voltage* between any two points. [3] Because EEG is the voltage difference between two sets of electrodes, the choice of reference scheme is critical. If the electrical fields between two locations is very similar, the voltage difference will be very small. If, however, there is a sudden change in the fields as you move across the cortex, this will produce a large voltage difference if measured by electrodes across this boundary. As such, if you are measuring spatially diffuse gradients in the electrical fields, you should choose a spatially distant and homogenous reference and if you want to measure spatially punctate gradients, you should choose a proximate reference.

For measuring large amplitude event-related signals reflecting the activity of large regions researchers frequently reference the EEG signals to either an average of all channels (average reference) or to a putatively neutral signal, such as that measured at the earlobes. For measuring high-frequency activity and single neuron responses from intracranial electrodes, researchers often choose a nearby electrode, or set of nearby electrodes, as the reference. Similarly, in clinical EEG studies used to localize epileptic spikes and seizures, neurologists also frequently use a bipolar reference scheme, referencing an electrode to its neighbor.

In the next several sections we briefly survey the major technologies for recording electrical signals in the human brain: Scalp EEG, magnetoencephalography, intracranial EEG, and neuronal recordings.

[3] Voltage is energy required to move a point charge between two locations. The reason it takes energy to move charge is because, as defined by Coulomb's law, charged particles exert a force on one another:

$$F_q = k_e \frac{q_1 \times q_2}{d_{1,2}^2},$$

where d is the distance between the charged particles, and k_e is the Coulomb constant. Thus, to move a point charge (define) from coordinates $\vec{x}_1$ to $\vec{x}_2$ requires energy (=force×distance), and we define this difference in potential energy (because the point charge is assumed to have no mass) as *voltage*.

Electro- and magneto-encephalography

To record scalp EEG signals, we apply electrodes to the scalp of the head, either individually or by fitting a cap with integrated electrodes to the head. After applying the cap, or the individual electrodes, one applies gel (or sometimes salt water) to each electrode to improve the conductance from the scalp to the contact. Many modern systems provide amplification of the electrical signal at the site of each recording contact, and then send these signals to a system that records the signals at a particular sampling rate (e.g., 2 kHz).

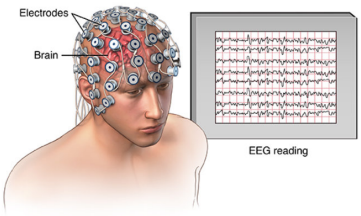

Figure 2.3: Illustration of a typical cap used for recording EEG signals, including an example of the EEG reading.

Magnetoencephalographic recordings require more complex equipment that measures the magnetic fields at the surface of the scalp. When ions move inside the brain they create currents, which induce magnetic fields that can be measured outside of the brain. These magnetic fields provide a view of currents that may be invisible to EEG measurements because currents produce magnetic fields that are orthogonal to the electrical fields. As such, by combining the measurement of magnetic and electrode fields, one can obtain a more complete view of the changing electrical signals in the brain. Researchers who record MEG (and EEG) signals often use algorithms to

attempt to localize the sources of these signals to specific locations in the brain. There is much controversy, however, regarding the accuracy of these localizations. This is because the inverse problem (identifying the internal signals that produce a particular distribution of external potentials) cannot be uniquely solved and as such, the algorithms make very strong assumptions that may be unwarranted.

Intracranial Electroencephalography

Intracranial EEG recordings, which are also sometimes referred to as electrocorticographic (or ECoG) recordings, involve the implantation of electrodes into the brain itself, either in the form of (subdural) grids placed directly on the cortical surface (see Fig. 2.4) or depth electrodes placed into the brain's parenchyma. Because they measure brain activity with high spatial and temporal resolution, surgically implanted electrodes help physicians diagnose and treat neurological conditions such as epilepsy, Parkinson's disease, and tumors. Here our focus is on ECoG recordings from patients undergoing invasive monitoring for drug-resistant epilepsy. In this procedure, surgeons implant ∼100–200 electrodes in widespread brain regions (Fig. 2.4A) to identify epileptic foci for potential surgical resection. Electrodes remain implanted throughout each patient's ∼1–3-week hospitalization. These electrodes include grid and strip electrodes (Fig. 2.4B,C), which record ECoG signals from the cortical surface, and depth electrodes (Fig. 2.4D), which penetrate the cortex to record field potentials from deep brain structures. In this text we will use the terms "ECoG" and "iEEG" interchangeably to refer to both surface and depth recordings. On occasion, surgeons implant microelectrodes, which record individual action potentials (Fig. 2.4E). We discuss these further below.

ECoG recordings measure brain activity directly with a resolution of $\sim 4\ mm^2$ (K. J. Miller, Sorensen, Ojemann, den Nijs, & Sporns, 2009). This high spatial resolution is a unique feature of ECoG compared to noninvasive methods like scalp electroencephalography (EEG) or magnetoencephalography (MEG). Noninvasive recordings, even with advanced localization algorithms, sometimes miss signals that are clearly visible with ECoG (Dalal et al., 2009). Furthermore, noninvasive techniques have difficulty isolating activity from deep brain structures and are relatively susceptible to muscle artifacts (Jerbi et al., 2009). Thus, ECoG is considered the clinical "gold standard" for accurately identifying seizure foci (Lachaux, Rudrauf, & Kahane, 2003; Crone, Sinai, & Korzeniewska, 2006). For the same reasons that ECoG recordings are useful to doctors, these data benefit researchers seeking to uncover the neural correlates of memory and cognition.

Each ECoG electrode measures the combined synaptic activity across the local population of neurons, rather than recording individual action potentials (Logothetis, 2003; Crone et al., 2006). Due to this aggregation, ECoG recordings measure the electrical activity that is synchronized across these neurons, which often includes rhythmic activity such as neuronal oscillations. Oscillations appear as periodic changes over time in the voltage observed from an electrode (Fig. 2.4F) and can appear at frequencies from $<$0.1 Hz to 500 Hz. Spectral methods can help researchers to identify oscillatory activity in the time series of recorded voltages. These methods, which we discuss in detail in Chapter 4, can be used to measure the amplitude of

the rhythmic component of a brain signal at different frequencies, as shown in the power spectrum in Figure 2.4G.

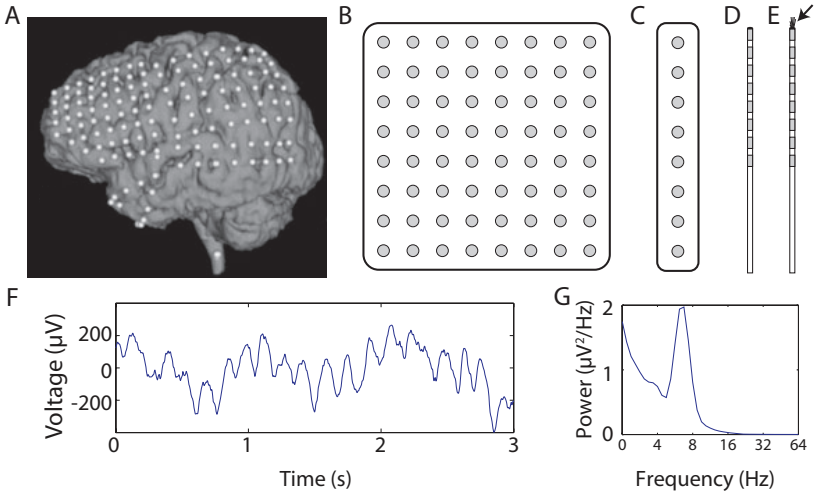

Figure 2.4: Intracranial EEG recordings. A. An MRI image of one patient's brain with the locations of implanted ECoG electrodes indicated with white dots. Modified, with permission, from (Towle et al., 2008). **B.** An illustration of an 8×8 electrode grid; gray shading indicates electrodes' conductive surfaces. (Illustrations not to scale). **C.** An illustration of an 8-electrode strip. **D.** A depth electrode with eight contacts. **E.** A depth electrode with microwires extending from the tip to record action potentials (marked by the arrow). **F.** A recording of ECoG activity from the right temporal gyrus. **G.** The power spectrum of the recording from Panel F, which shows that this trace exhibits a robust theta oscillation.

Microelectrode recordings

The most common method to measure *in vivo* neuronal activity is through extra-cellular recordings. In this method, a small micro-wire ($< 60\mu m$) is manually placed in the immediate vicinity of an intact neuron (Moxon & Nicolelis, 1999). Using single tungsten electrodes, micro-wire arrays and drivable tetrodes, extracellular neuronal recordings have long been a staple of animal electrophysiology studies. Indeed, if these electrodes are placed close enough to a neuron, activity from a single cell can be obtained (single-unit activity; SUA). Failing this, it remains possible to record multi-unit activity (MUA) from the collection of cells in a somewhat larger neighborhood around the recording site (Buzsáki, 2004; Gold, Henze, Koch, & Buzsaki, 2006). Both SUA and MUA recordings have contributed substantially to our understanding of the neural origin of cognitive functions.

Because single unit recordings have yet to play a clinical role in the in the treatment of epilepsy, only a handful of research-oriented clinical centers obtain these recordings from patients. Indeed, there are substantial technical challenges to obtaining quality unit recordings in the clinic, and although many of these challenges can be surmounted (see Hefft et al. (2013)), this type of research is not for the faint-of-heart.

To ethically obtain extracellular neuronal recordings in the clinic, one needs to use specialized depth electrodes such as the commercially available, FDA-approved combined macro-micro depth electrodes designed by Itzhak Fried's team at UCLA (Behnke-Fried depth electrode and inner-wire bundle; Ad-Tech Medical, Racine WI) (Fried et al., 1999; Babb, Carr, & Crandall, 1973). This electrode is manufactured as two separate components. The clinical component (henceforth referred to as the Behnke-Fried depth; BFD) consists of eight standard cylindrical, depth macro-electrodes (90% platinum, 10% iridium alloy contacts, 1.3 mm in diameter, 0.8 mm in length)

that are embedded on the surface of a silastic tube with a hollow lumen. The research component (henceforth referred to as the inner-wire bundle; IWB) runs through the lumen of the BFD and consists of eight micro-wires (platinum-iridium; 40 μm diameter; Teflon insulation) and one reference wire (an additional micro-wire stripped of insulation).

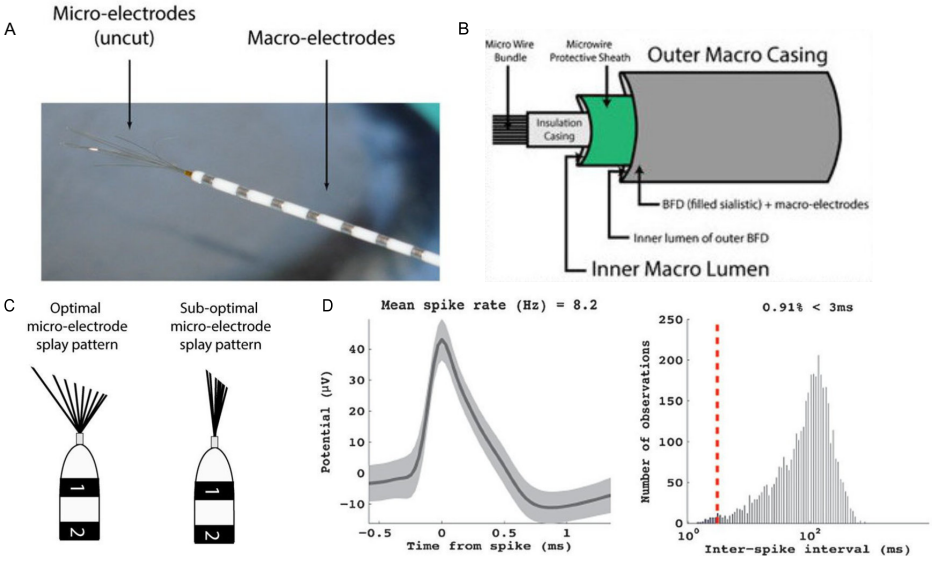

Figure 2.5: A. Projections of the micro-wires out of the end of the macro-electrode as received by the manufacturer (AdTech macro-micro electrode). **B.** Schematic of the outer and inner casing included in the macro-micro electrode. **C.** Cartoon example of optimal and sub-optimal micro-electrode splay patterns and staggered micro-wire lengths. **D.** The average waveform (left panel) and the distribution of inter-spike intervals (right panel) are shown.

One of the main challenges to obtaining quality unit recordings is due to the fragile nature of the microwires. Indeed, a significant percentage can break at some point during the patients hospitalization. For this reason it is essential to conduct impedance testing at various points post implantation. If the broken microwire is the reference channel, one can re-reference to a different microwire, insuring that the chosen reference is not recording any neuronal spikes. For the Behnke-Fried electrodes, typical impedances of intact electrodes range between 50 and 500 kQ (Hefft et al., 2013). If one obtains measurements greater than 1 MQ one should assume that there is a failure in the wire. By taking adequate precautions to insure against broken electrodes one should obtain a yield of 2-3 neurons per depth electrode.

A number of software packages have been developed to help researchers isolate extra-cellular single unit activity from high-frequency microwire recordings (typically one uses a sampling rate in excess of 20 kHz for such analyses). One popular package that is used for these analyses is called WaveClus (Quiroga, Nadasdy, & Ben-Shaul, 2004). The output of this package is a series of spike times. It is often helpful to convert these into a continuous measure of the firing rate of each cell, and this can be easily done using a variety of filtering tools (e.g., applying a Gaussian filter to the data).

Why electrophysiology?

Supposing that our objective is to understand how memory works. Why study the brain? Many would answer by rejecting the premise behind this question. If the brain enables us to perform complex behaviors it is inher-

ently interesting and we should try to understand it. Moreover, in cases of neurological and psychiatric diseases, understanding the brain has helped to advance diagnosis and treatment.

Nonetheless, it would be valuable to consider how understanding electrophysiology can directly advance our understanding of cognition itself.

In answering this question it is helpful to consider how we might understand cognition without electrophysiology. Here, the answer provided by traditional cognitive psychology is that we learn about the mind by studying behavior, and specifically we study the two primary dependent variables recorded during a memory task: accuracy and response time. Even cognitive psychologists, however, have found it very helpful to supplant these primary measures with additional behavioral measures whenever possible. For example, in the study of recall, the order of responses has helped elucidate the mechanisms of retrieval. In the study of recognition memory, researchers often ask subjects to make confidence judgments ("How sure are you that the item was on the list"), or to judge the subjective quality of remembering ("Did you recollect specific details of the item's encoding or did it just feel familiar"). Whenever possible, cognitive psychologists have embraced more detailed information about the process of remembering and used that information to test theories of memory. One could reasonably argue that electrophysiological recordings should serve this role as well, providing millisecond resolution data related to the processes of encoding, retention and retrieval. But to serve this role, researchers need to understand how to analyze and interpret the electrophysiological data.

As an example, consider the distinction between encoding and retrieval processes. During the study phase of an experiment we normally ask a subject to learn a list of memoranda (e.g., words or pictures) for a later test. There is no overt behavior measured during this stage, yet we know that encoding efficiency varies across both items and lists (e.g., Kahana, Dolan, Sauder, & Wingfield, 2005; Kahana & Aggarwal, 2018). Without easy access to physiological measurements, cognitive psychologists developed elaborate behavioral methods to help uncover variability in encoding processes (e.g., Rundus, 1971). But by recording physiological activity during encoding, and relating that activity to subsequent memory performance, we can both identify those neural features measured during encoding that predict subsequent memory, and we can use those neural features to help characterize the previously unobserved variability in encoding processes.

An arguably even more powerful use of electrophysiological recordings would be to test theories of memory, but for this to happen, the theories need to make specific neural predictions. In some cases, neurobiologists studying brain recordings in animal models posit connections to memory phenomena that are not easily measured in animals. Chapter 2 described several such models, including the Lisman-Idiart-Jensen model of memory scanning. In that model, two brain oscillations work together to support *multiplexing* of short-term memories: enabling a small number of neural representations (of items, presumably) to repeat in serial order at the frequency of the theta rhythm. Cognitive neuroscientists have used both electrophysiological and haemodynamic (i.e., fMRI) recordings, in humans, to test the predictions of this model.

Mathematical models of human memory, developed to explain behavioral rather than neural data, frequently make predictions about the dynamics

of internal representations. These internal representations had long been unobservable, but with modern recording techniques we can observe high-dimensional patterns of neural activity and study their dynamics during the course of a memory experiment. These studies have allowed us to identify putative representations in the brain that exhibit dynamics that match predictions from cognitive theories. Although it is hard to rule out cognitive theories with neural data (because you don't observe everything in the brain) it is nonetheless valuable to know which cognitive theories make predictions that can be related to neural processes. The success of one account over its competitor may reasonably lead to an increased investment in further experiments aimed at testing the more successful theory.

In his influential analysis of the visual system, David Marr suggested a distinction between three levels of analysis of a complex system: the computational, the algorithmic and the implementational. Most complex systems evolved to solve a problem, to compute something. Work at the computational level of analysis seeks to identify the problem that a system is trying to solve. The memory system, for example, needs to be able to deploy the relevant information from the past to solve the decision problems of the present. Much like a computer uses algorithms to solve any computational problem, the brain likely has its own algorithms, to solve its computational problems. The development of memory models described in Chapter 1 illustrate the algorithmic approach to studying memory. Finally, the brain uses the nervous system to implement these algorithms. Viewed through the lens of Marr's analysis, electrophysiological data serve as an important bridge between the algorithmic level of cognitive theories of memory, and the implementational level of neural recordings. Having bridges between levels of analysis is essential if we are to someday understand how the brain gives rise to complex cognition and behavior.

Figure 2.6: David Marr (1945 – 1980) was a cognitive scientist at MIT known for his models of primate visual perception.

Neurons, Fields and Networks

The cognitive operations that contribute to episodic memory are instantiated in the patterns of activity produced by networks of neurons distributed throughout the brain. When neurons are active they generate *action potentials*, electrical events that alter the electric field in the neuron's extracellular environment. The action potentials of single neurons can be recorded using microelectrodes; larger electrodes implanted intracranially or non-invasively placed on the surface of the scalp can also be used to record changes in the aggregate electric field caused by the summed activity of many hundreds or thousands of neurons. Measuring these electric field changes is the basis of using intracranial and scalp electroencephalography to study memory function (see Chapter 3). Here, we first review some basic physical principles concerning the propagation of electrical charge that underlie the signals captured by these recording methods.

Basics of neurophysiology

Neurons

Structure Although there are many types of neurons in the nervous system, they share some common morphological characteristics. Neurons are made up of a cell body (*soma*) that contains the cell nucleus and performs the metabolic functions required by the rest of the cell. There are two types of processes that extend out of the soma and that enable communication between neurons: several *dendrites* and a single *axon*. The dendrites are short and branch out from the soma like trees; their role is to receive input signals from other neurons. The axon is generally much longer and is the neuron's output structure for sending signals to other neurons. The axon of a neuron extends out and terminates near the dendrites of a single or many other neurons. The axon of one neuron is separated from the dendrite(s) of another neuron by a small gap called the *synapse*.

Ion channels Action potentials are initiated at the soma in a region called the axon hillock and then travel down the axon via passive or saltatory conduction (described below). Communication between neurons depends on fast propagation of electrical signals witin a neuron from the axon hillock to the synapse, followed by slower chemical communication via the release of neurotransmitters into the synapse. The rapid propagation of the action potential is accomplished by the flow of ions into and out of the cell through ion channels that span the cell membrane between the intracellular and extracellular media. Ion channels consist of proteitns that open and close in response to specific electrical or chemical signals. The flow across ion channels occurs very rapidly, leading to large currents and changes in voltage across the cell membrane (Equation 2.1), compared to the resting state.

Resting membrane potential At rest (i.e. when a neuron is not in the middle of generating an action potential), there is a voltage difference across the cell membrane, meaning the difference in electric potential is non-zero between the inside and outside of the neuron [4]:

$$V_m = V_{intra} - V_{extra} \tag{2.2}$$

where V_x are the intra- and extracellular voltages. When V_m is measured in neurons at rest it is referred to as the *resting membrane potential* and is typically between -60 mV to -70 mV. Electrical signaling occurs when V_m deviates from its resting value due to opening and closing of ion channels that cause current to flow across the membrane. Deviations that cause V_m to become less negative (decreasing the voltage difference across the cell membrane) are called *depolarizations* and deviations that cause V_m to become more negative are called *hyperpolarizations*.

The resting membrane potential is important because it means that once voltage-gated ion channels open in response to depolarization inside the cell, the resting imbalance in voltage between the inside and outside of the cell will cause positively charged ions to flow into the cell. The flow of charged particles is passive, meaning that the cell does not expend energy to move ions across the cell membrane. The flow of charged particles also creates a change in the extracellular voltage that can be measured with microelectrode recordings.

Action potentials and synaptic transmission When the membrane potential of a neuron reaches a threshold of depolarization at the axon hillock,

[4] Current is the flow of electric charge through a medium, typically carried by electrons or ions. In the case of a neuron, current is induced when ions flow through voltage-gated channels as part of action potential propagation (see *Neurons*). The strength of current flow is dependent on the voltage difference between two points, where voltage refers to the difference in electric potential energy between two points. In a conductor, current and voltage are related directly through *Ohm's Law*

$$V = IR \tag{2.1}$$

which indicates that the amount of current that is present in a medium is a function of the voltage difference (V) between the two points and the resistance (R) of the medium. One can think of an analogy to the flow of water through a pipe. If the pipe is of small diameter (high resistance) then it will be difficult for much water to flow through the pipe, even if it is oriented at or close to vertical (high voltage). In contrast, the same vertical pipe orientation (voltage) could produce a much larger flow of water (current) in a pipe of larger diameter (low resistance).

it triggers an action potential. The threshold-crossing depolarization leads to the opening of voltage-gated Na^+ channels in the area of the initial depolarization. The opened channels allow excess Na^+ outside the cell to flow into the cell, leading to further depolarization, which in turn causes the opening of additional voltage-gated channels nearby. This positive feedback loop occurs in such a way that Na^+ channels are sequentially opened along the length of the axon. The resultant propagation of the depolarization leads to current flow along the length of the axon in the extracellular space.

As the depolarization moves down the length of the axon, slower cellular processes begin to take effect in which K^+ ions flow out of the cell, leading to hyperpolarization and restoration of V_m back to the resting membrane potential. When the action potential propagates to the end of the axon, it causes the release of chemicals called *neurotransmitters* into the synapse between the end of the pre-synaptic axon and the post-synaptic dendrites of other cells. The released neurotransmitters can then bind to receptors in the post-synaptic neuron(s) and then, depending on whether the synapse is excitatory or inhibitory, cause either hyper- or depolarization of the post-synaptic cell. If the synapse is excitatory, the pre-synaptic action potential may thus evoke an action potential in the post-synaptic neuron.

Ensembles and their fields

The extracellular field Above, we described how action potentials alter the extracellular field during communication events between neurons. Here, we review other sources of current flow that alter the extracellular field, and how these lead to voltage fluctuations that are detectable with electrophysiological recording methods.

The extracellular field at any point is the sum or superposition of many types of ionic events occurring in the area near the recording electrode. When recording the extracellular field using scalp electroencephalography (EEG) or intracranial EEG (see Chapter 3), many small currents need to overlap in space and time in order to be detectable. For this reason it is generally accepted that the largest contributor to the extracellular field is synaptic activity because of the way that dendrites (and their synapses) cluster near the cell body. This allows the currents generated by the opening of many ion channels within a local area to aggregate, generating a *sink* at the point at which positive ions flow into the cell. In order to maintain conservation of charge, the sink generates a corresponding *source* and together the sink and source are referred to as a *dipole*. The cumulative effect of many dipoles in close spatial proximity contributes to measurable voltage fluctuations in the extracellular medium, and decays with distance (r) as $1/r^2$.

An important factor that impacts the strength of fluctuation in the extracellular field is the geometry of the dipole-generating neural elements. In general neurons in the cortex of the brain are arranged in layers such that the apical dendrites of neighboring pyramidal cells are oriented parallel to each other. This parallel organization is conducive to the superposition of neighboring dipoles from many active neurons, which leads typically to large signal fluctuations in cortical recordings. The other influence on the strength of the extracellular field is the temporal synchrony of the activity, which is necessary in order for the induced electrical events to superimpose.

Oscillations in the local-field potential

Because the neural events that underlie fluctuation in the extracellular field must show some degree of temporal synchrony in order to be detectable using scalp and intracranial EEG, there is has been significant effort devoted to understanding several prominent oscillations that are observed in the EEG recordings of animals and humans. Here we review some of these signals and their relation to cognitive processes.

Theta oscillations, place cells, and spatial memory

Because the hippocampal formation plays a critical role in learning and memory (see Chapter 2) researchers have long sought to understand how patterns of electrical activity measured in the rat hippocampus relate to animal behavior and cognition. Technological advances in the 1960s enabled researchers to measure these patterns of electrical activity in awake, behaving animals (e.g. Vanderwolf, 1969). One such pattern of electrical activity is the hippocampal theta rhythm—a 4-10 Hz oscillation that appears when an animal is alert and interested in its surroundings (Bland, 1986). Although early work emphasized the relation of theta specifically to motor activity, subsequent studies demonstrated a broader role for theta in the hippocampal computations underlying spatial learning (Kahana, Seelig, & Madsen, 2001).

Theta's functional importance derives from several lines of evidence. First, theta appears to act as a windowing mechanism for synaptic plasticity, with synapses being strengthened when pre- and post-synaptic neurons fire at the peak of the theta rhythm, and synapses being weakened when neurons first at the trough of the theta rhythm (Huerta & Lisman, 1993; Hölscher, Anwyl, & Rowan, 1997).

If the phase of theta is crucial for experience-dependent changes in synaptic strength, then one might expect that stimulus events would produce a reset or phase shift in ongoing theta. Consistent with this hypothesis, the activity of hippocampal theta appears to be phase locked to stimuli when an animal is incentivized to maintain the stimulus information in memory (B. Givens, 1996). Taken together, these observations help to explain how important sensory input undergoes neural encoding (Hasselmo, Bodelon, & Wyble, 2002).

Second, theta's functional importance has been demonstrated through attempts to block theta. Theta can be blocked by lesioning a region in the brain known as the medial septum. Such lesions, in addition to blocking theta, produce severe impairments in memory function (e.g. see B. S. Givens and Olton (1990)). Although neither prior learning of spatial information nor hippocampal place representations are impaired by septal lesions, such lesions do impair the acquisition of new spatial information (Leutgeb & Mizumori, 1999). This evidence suggests that theta has a role in memory, but it is difficult to dissect the specific effect on theta from the concomitant cholinergic loss. Adaptive electric-field feedback, which can accentuate or minimize a specific frequency band in generated field potentials (Gluckman, Nguyen, Weinstein, & Schiff, 2001), may be used to directly assess the effects of theta manipulation in the rat.

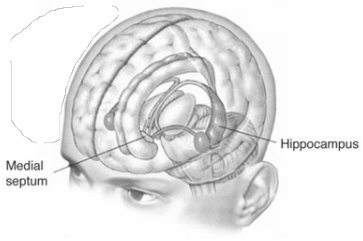

Figure 2.7: Illustration of deep brain structures including the medial septum and hippocampus.

Human Theta

By the 1990s, theta was extremely well described in rodent studies, yet there was scant evidence for any homologous rhythm in primates. Electrophysiological investigations in monkeys focused primarily on sensory and motor processes and their neural correlates, and no documented evidence for task-dependent theta in monkeys had been reported prior to 1999. Motivated by studies of theta activity during rodent navigation, Kahana, Sekuler, Caplan, Kirschen, and Madsen (1999) had patients with subdural grid electrodes learn to navigate through virtual, three-dimensional rendered mazes. Their recordings revealed clear oscillations in the unfiltered iEEG traces and demonstrated that, during maze navigation, intermittent bouts of theta activity appear with greater probability during longer mazes, even when controlling for degree of mastery.

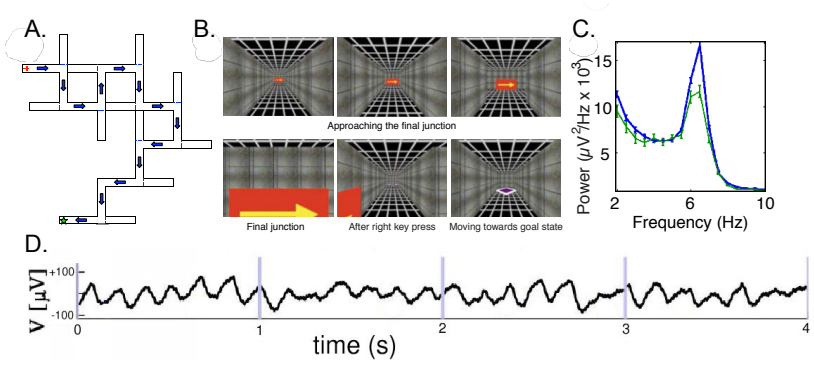

Figure 2.8: Theta oscillations during maze learning. A. Blueprint of a sample maze. Participants navigated T-junction mazes, from a starting point (+) to a goal position (⋆). **B.** Sample views during study phase. During the first 4 traversals, arrows denoted the correct path (*study*). Participants then repeatedly traversed the maze without arrows (*test*) until they traversed it 3 times consecutively without errors. Maze length was varied (short mazes had 6 junctions; long mazes had 12 junctions). **C.** Power spectra at an electrode in the inferior frontal gyrus indicating greater ∼6.5-Hz theta power for correct traversals of long mazes (thick blue) compared with traversals of short mazes (thin green). This effect was observed at multiple sites in each of 5 patients (Kahana et al., 1999; Caplan et al., 2001). **D.** Theta oscillations revealed by intracranial EEG recorded from an electrode on the inferior frontal gyrus during a virtual maze navigation task.

Caplan et al. (2001) showed that the effect of maze length on theta does not reflect the increased difficulty of encoding or retrieval at individual choice points. Rather, it reflects a global difference between long and short mazes. Caplan et al. found that gamma activity, but not theta, increased with increasing difficulty of individual choices at maze junctions. It was at these difficult junctions in which the learning requirements would have been greatest.

Because subjects often learn T-mazes as a verbal sequence of left and right turns (Kirschen, Kahana, Sekuler, & Burack, 2000), it is difficult to make strong inferences about the role of theta in spatial processing from Kahana and Caplan's earlier studies. To help address this confound, we developed a task called "Yellow Cab" in which subjects play the role of a taxi driver, driving through a virtual town in search of passengers and delivering those passengers to their requested destinations (see Fig. 2.9a,b). Over repeated deliveries, passengers learn to find the shortest path between the random locations in which they find their passengers and the fixed locations of landmarks within the environment (Caplan et al., 2003; E. L. Newman et al., 2007).

Caplan et al. (2003) found increased theta activity at widespread cortical sites when subjects were moving (compared with periods in which they were still). Extending these results by comparing iEEG recordings from hippocampus and neocortex, Ekstrom et al. (2005) found increased theta activity in both regions when subjects were moving within the virtual town.

Watrous, Fried, and Ekstrom (2011) further demonstrated that the frequency of human hippocampal theta increases with a subject's virtual speed, as in rodents. Finally, the amplitude of theta appears to positively correlate with performance in a virtual navigation task (Cornwell, Johnson, Holroyd, Carver, & Grillon, 2008).

The high amplitude theta activity reported during human maze learning appears very much like the theta seen in rodents during spatial exploration. Following the initial reports of task-dependent human theta, some investigators suggested that these observations may be specific to tasks that involve a spatial component (O'Keefe & Burgess, 1999); however, the discovery of rodent theta during non-spatial learning tasks (Hasselmo et al., 2002; Ekstrom, Meltzer, McNaughton, & Barnes, 2001) and numerous findings of task-dependent theta in non-spatial memory paradigms indicates that theta plays a far more general role in human cognition. Subsequent chapters discuss the extensive literature demonstrating the importance of theta-frequency activity in a wide range of cognitive tasks.

The ability of human intracranial recordings to record from deep brain structures allowed researchers to compare the properties of hippocampal theta between humans and animals. As a result, researchers identified two important interspecies differences. In rodents, hippocampal theta oscillations reliably appear at 4–8 Hz (Buzsáki, 2005). However, in humans, hippocampal oscillations usually appear instead at 1–4 Hz (Bódizs et al., 2001; de Araujo, Baffa, & Wakai, 2002; Ekstrom et al., 2005; Jacobs, Kahana, Ekstrom, & Fried, 2007; Babiloni et al., 2008; Cornwell et al., 2008; Clemens et al., 2009). Furthermore, whereas rodent theta oscillations are routinely sustained for over ten seconds (O'Keefe & Recce, 1993; Buzsáki, 2005), human hippocampal oscillations usually appear only transiently (Caplan et al., 2003; Ekstrom et al., 2005) and sometimes not at all (Niedermeyer, 2008). Despite these differences, it seems that human 1–4-Hz hippocampal oscillations are functionally analogous to rodent 4–8-Hz theta. During navigation, both of these oscillations increase in amplitude during movement (Ekstrom et al., 2005; Buzsáki, 2005; Jacobs et al., 2010) and the phase of both oscillations modulates neuronal spiking (O'Keefe & Recce, 1993; Klausberger et al., 2003; Jacobs et al., 2007).

A neural GPS system

Advances in the ability to record extracellular action potentials from multiple hippocampal neurons while rodents performed complex navigation-based learning tasks led to the discovery of the neurons in the rodent hippocampus that represent the animals' spatial location within a given environmental (O'Keefe & Dostrovsky, 1971). These so-called *place cells* fire preferentially whenever the animal passes through a region of their environment, termed the *place field*. When navigating through an environment with an open layout, place cells fire without regard to the direction of movement, exhibiting the property of *omnidirectionality*; in contrast, place cells generally exhibit a preferred direction when the environmental layout consists of narrow paths that the animal would typically traverse in one of two directions, as in a T-maze. In the rodent hippocampus, approximately 25% of all pyramidal cells appear to represent information about the animals position within their environment.

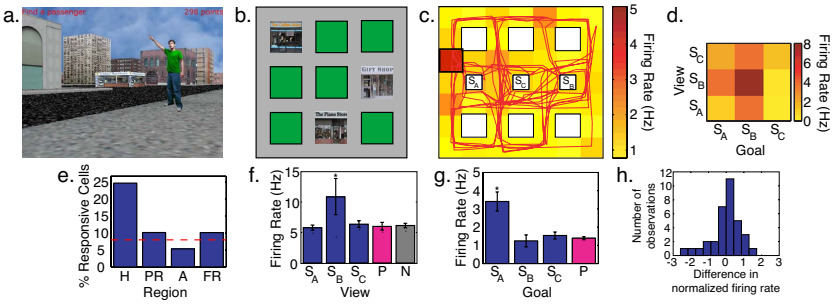

Figure 2.9: Cellular responses during virtual spatial navigation. a. View from within Yellow Cab. **b.** Map illustrating a possible town layout with three stores. **c.** Firing rate of a hippocampal neuron overlaid on a town map. The participant's path is shown in red. **d.** A view–goal conjunction cell that responded preferentially when the goal store (S_B) was also in view. **e.** Proportion of place cells ($p < .01$) in hippocampus (H), parahippocampal cortex (PR), amygdala (A), and frontal regions (FR). Red line indicates bootstrap-determined false-positive rate. **f.** A view cell in parahippocampal cortex that responded preferentially when viewing store B (S_B). Also depicted are cellular responses to views of a passenger (P) and a set of neutral views (N) of neither stores nor passengers. **g.** A goal cell in the amygdala that responded preferentially when seeking store A (S_A). **h.** Regions of high firing included high numbers of traversals in different directions. The distribution of firing-rate differences across these traversals was centered on zero, indicating that the cells responded in an omnidirectional manner.

The hippocampal theta rhythm appears to also play an important role in the neural coding of place. As a rat traverses a place field, hippocampal place cells fire at a progressively earlier phase of the ongoing theta oscillation (O'Keefe & Recce, 1993; Skaggs, McNaughton, Wilson, & Barnes, 1996). This information significantly improves accuracy in reconstructing the animal's position in space (Jensen & Lisman, 2000), providing additional support for the hypothesis that the phase of theta at which cells fire plays an important role in the coding of place information in the rat hippocampus. Phase precession may also provide a more general mechanism for sequence learning in episodic memory outside of purely spatial domains (Buzsáki & Moser, 2013).

Ekstrom et al. (2003) asked whether neurons in the human hippocampus similarly encoding spatial information in a context-dependent manner. Using a variant of the Yellow-Cab taxi-driver game described above, Ekstrom et al examined the firing patterns of 378 neurons recorded across a sample of 7 neurosurgical patients. Figure 2.9c shows an example of a hippocampal cell whose firing rate peaked at its *place field* in a northwest region of the virtual town. Ekstrom et al observed significant place-selective neural activity in 39 cells, with the majority of place-responsive cells being found in the hippocampus (Fig. 2.9e). Ekstrom et al further observed that the majority of these cells responded in a direction-independent manner (Fig. 2.9h) mirroring the finding of omnidirectional place cells in the rat hippocampus (O'Keefe & Dostrovsky, 1971).

An advantage of virtual navigation tasks is that software is able to track which landmarks subjects could view during each frame in the game. Ekstrom et al used these data to determine how neurons responded to visual information, such as whether a target store was in view. Of theoretical interest was whether view-responsive cells exist in the same medial temporal lobe (MTL) brain network that appears to be involved in navigation. This analysis identified 41 cells that responded preferentially when participants viewed specific landmarks (Fig. 2.9f). Ekstrom et al observed a highly significant dissociation between the anatomical distributions of place and view cells, with place cells being most prevalent in the hippocampus and view cells being most prevalent in the parahippocampal region. This is consistent with a considerable body of functional neuroimaging evidence implicating the parahippocampal region in the processing of spatial scenes (Epstein & Kanwisher, 1998; Epstein, 2005).

In addition to studying the cellular correlates of place and view, the Yellow Cab paradigm provided the opportunity to explore the neural repre-

sentation of goals. Because participants were asked to drive to one of three randomly chosen goal destinations on each trial, we were able to determine whether individual neurons were responsive to specific goal destinations independently of place or view. This analysis revealed 69 goal-responsive cells (Fig. 2.9g). Consistent with evidence from recent evidence in rodents, Ekstrom et al found these cells in frontal brain regions (Hok et al., 2005).

Finally, Ekstrom et al examined whether cells responded conjunctions of these variables. For example, one might expect that some goal cells would respond most strongly when the goal destination comes into view. This can be seen in a right amygdalar cell that responded preferentially to a view of a given store when that store was the participant's intended destination (Fig. 2.9d). The authors observed significant numbers of neurons that responded to conjunctions of place and view, and to conjunctions of place and goal. The existence of these conjunctive cells was anticipated by Burgess, Becker, King, and O'Keefe's (2001) computational model of spatial navigation.

Building on the discovery of hippocampal place cells, which identify unique locations within a specific environment, a team led by husband and wife neurobiologists, Edvard and May-Britt Moser, discovered an amazing class of neurons, termed *grid cells* that fire when a rodent traverses locations corresponding to the vertices of a triangular grid spanning the animal's environment. Unlike place cells, which reside primarily in the hippocampus, grid cells appear most prominently in the medial entorhinal cortex. By combining place and grid responses, the central nervous system is able to precisely represent an animal's location within an environment. Recording from neurosurgical patients as they performed a virtual navigation task in an open-field environment, Jacobs et al. (2013) reported grid-cell like responses in the human entorhinal cortex, and demonstrated that these cells exhibit similar response profiles as those reported in rodents. [5]

Place and grid cells exist as part of a broader network of neurons that encode other navigationally relevant variables including those related to local geometry (border cells), orientation (head direction cells), current navigational goals, and scene information. Current theories of spatial memory suggest that the hippocampus constructs a viewpoint-independent (allocentric) representation of the current environment based on not only place and grid cells, but also on inputs from parahippocampal cortex (scene information), parietal cortex (viewpoint-dependent egocentric information), temporal cortex (featural representations of environmental stimuli), and prefrontal cortex (goal- and route-planning information). The major challenge for the field is understanding how the brain uses these representations to navigate, reason about spatial relations, and flexibly distinguish between and exploit different spatial representations in a context-dependent manner. We take up some of these issues in a later chapter.

[5] John O'Keefe, Edvard Moser, and May-Britt Moser received the 2014 Nobel Memorial Prize for their characterization of the cellular networks underlying spatial cognition. This award followed the discovery of similar neural responses in numerous other species, including humans.

Do rhythms help to organize spikes?

Research in animals shows that brain oscillations provide a neuronal timing signal that allows neurons to encode information by spiking at a particular phase of an oscillation—a phenomenon called phase coding (O'Keefe & Recce, 1993; Fries, Nikolić, & Singer, 2007). To examine the prevalence and properties of phase coding in humans, Jacobs et al. (2007) examined how

neurons in widespread regions varied their instantaneous firing rate according to the phase of ongoing oscillations. This work found that many neurons were *phase locked* to oscillations, a phenomenon in which they increased their firing rate at a particular phase of these oscillations. Figure 2.10A shows the activity of a neuron that exhibits this phenomenon by spiking just before the peak of the theta oscillation. The properties of neuronal phase locking varied between high- and low-frequency oscillations. Neurons phase locked to oscillations at frequencies slower than 10 Hz had various preferred phases, whereas neurons phase locked to oscillations faster than 10 Hz had preferred phases near the oscillation's trough. This indicates that oscillations faster than ∼10 Hz reveal specific times (the trough of the oscillation) when many neurons are active, whereas slower oscillations cannot predict population spike times with this level of precision.

In addition to examining the timing of individual action potentials, a different set of studies examined the relation between the rate of neuronal spiking and the amplitude of oscillatory activity. In some cases, neuronal firing rate is well predicted by the amplitude of simultaneous oscillations (Fig. 2.10B). However, the details of this relation dramatically vary according to the oscillation and brain region being examined. Oscillations at high frequencies (>10 Hz) in sensory cortex correlate positively with neuronal spiking (Nir et al., 2007) and a similar, but weaker, pattern appears in hippocampus (Ekstrom, Suthana, Millett, Fried, & Bookheimer, 2009). In contrast, low-frequency oscillations exhibit varied correlations with single-neuron spiking: In neocortex, theta- and alpha-band oscillatory power is negatively correlated with neuronal spiking (Nir et al., 2007), but in hippocampus these oscillations do not correlate with spiking rate (Ekstrom et al., 2009).

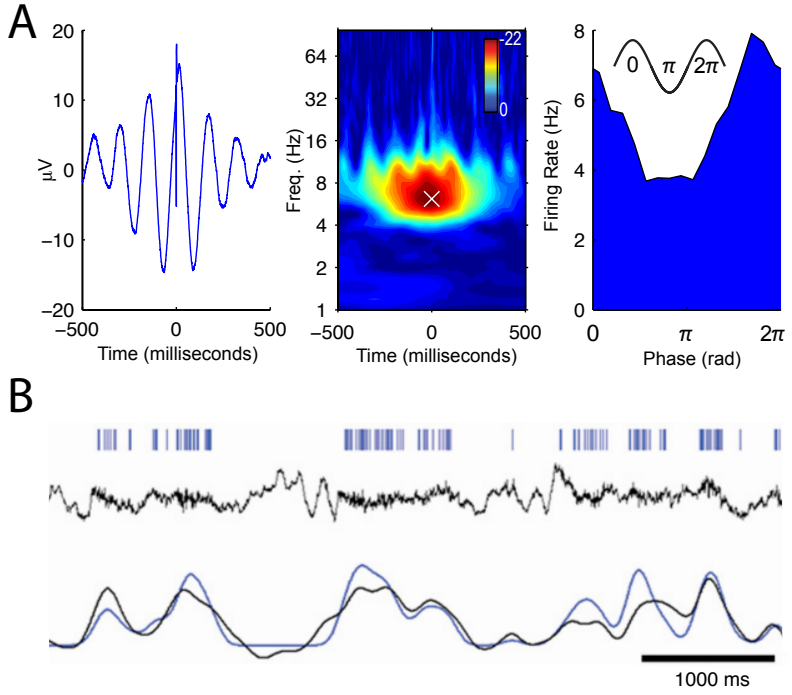

Figure 2.10: The relation between oscillatory brain activity and neuronal spiking. A. Activity of a neuron in the right superior temporal gyrus that spiked just before the peak of the theta oscillation. Left panel, average local-field potential (LFP) computed relative to each spike. Middle panel, z-score index of the phase uniformity at the time of each spike, as a function of frequency and time offset. White '×' indicates frequency of peak phase locking. Right panel, firing rate of this cell as a function of instantaneous theta phase at the frequency of peak phase locking. Adapted from (Jacobs et al., 2007). **B.** The activity of a neuron in auditory cortex whose spiking was tightly coupled to the amplitude of simultaneous gamma oscillations ($r = 0.84$). Ticks in top row indicate action potentials. Middle row depicts the LFP signal filtered to only include frequencies below 130 Hz. Bottom row shows the correlation between LFP gamma power (black) and neuronal firing rate (blue). Adapted from (Nir et al., 2007).

"Broadband" EEG fluctuations as a marker of neural activity

Reports of strong correlations between neuronal firing and narrowband activity (i.e. oscillations) have supported the view that oscillations reflect synchronized spike timing in large neuronal ensembles (Singer & Gray, 1995; Logothetis, 2003; Fries et al., 2007). This follows in part from the temporal-binding hypothesis (von der Malsburg, 1981), which proposes that synchronized neural activity can solve the "binding problem" by linking multiple neuronal signals (Köhler, 1947; Koffka, 1935; Kanisza, 1979; Pal & Pal, 1993). However, an emerging body of research has shown that apparent correlations between spikes and gamma-band LFP activity are actually due to broadband LFP patterns, rather than band-specific oscillations (e.g., K. J. Miller et al., 2014; Burke, Ramayya, & Kahana, 2015).[6]

Using simultaneously recorded LFP and single-neuron activity, Manning, Jacobs, Fried, and Kahana (2009) investigated the relationship between broadband activity, narrowband oscillatory activity, and underlying neuronal spiking. Consistent with previous studies, they found a population of *narrowband-shift neurons*, which varied their firing in proportion to LFP power at specific frequency bands. Narrowband-shift neurons were present throughout the brain, but were especially prevalent in the frontal cortex and amygdala. In addition, they observed a larger population of *broadband-shift neurons*, which varied their firing with the overall height of the LFP power spectrum at all frequencies. Broadband-shift neurons appeared in all examined brain regions, but were especially prevalent in the medial-temporal lobe. Broadband increases in LFP power were almost exclusively positively correlated with single-neuron firing, providing a robust estimate of neuronal firing. Below we describe this study in greater detail, primarily to illustrate the methods used in the analysis of both spectral EEG activity and neuronal spiking.

[6] The $1/f$ spectrum (or pink noise spectrum) is instantiated in many systems in the natural world and refers to a characteristic of a signal in which the power spectral density of the signal is inversely proportional to frequency. $1/f$ noise has been observed in the flow of traffic on highways (Musha & Higuchi, 1976), the structure of DNA base sequences (Voss, 1992), and in the timeseries of errors that people make in perceptual memory tasks (Gilden, Thornton, & Mallon, 1995). Neural activity recorded using electrophysiological measures also follows a $1/f$ pattern, meaning that power at low frequencies is much higher than at low frequencies (also see Chapter 5).

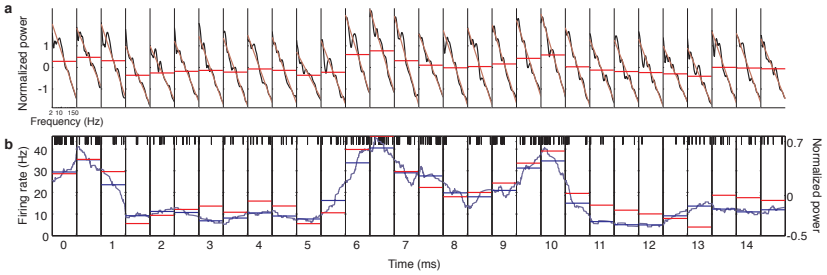

Figure 2.11: LFP power and neuronal firing time series. Each box details the activity in one 500-ms epoch. **a.** This panel illustrates how various features of the LFP change over time. In each epoch, the black lines indicate the overall LFP power spectrum, brown lines indicate robust-fit lines, and the horizontal red lines indicate mean broadband powers. **b.** This panel illustrates changes in neuronal firing rate concurrent with changes in the LFP power spectrum. Black vertical ticks represent the times when individual spikes occured, dark blue lines indicates the smoothed firing rate (see *Methods*), and horizontal blue lines indicate mean firing rates in each epoch. Mean broadband power is shown in panel b (horizontal red lines) on a different scale (indicated at right).

Using recordings of 2,030 neurons from 20 neurosurgical patients (Manning et al., 2009) determined how moment-to-moment variations in the local field potential related to simultaneous changes in the firing rates of nearby neurons. Spectral analysis methods provide a tool for transforming the time series of voltage activity (the EEG or local-field potential) into distinct frequency components. Even if there are no oscillations present in the neural activity, the spectrum will still have energy at varying frequencies. For a randomly varying voltage signal that is based, in part, on its prior state (i.e., an autocorrelated signal), the power spectrum will have more energy at lower frequencies than at higher frequencies, with power falling off as

a function of the reciprocal of frequency, often raised to an exponent (i.e., Power$(f) \sim 1/f^{\alpha}$). An increase in overall variability in the signal would raise the power at all frequencies, conforming to the overall shape of the colored noise distribution. By contrast, a narrowband oscillation would appear as a peak in the power spectrum, rising above the background ($1/f^{\alpha}$) distribution.

Figure 3.11 illustrates the relation between broadband power shifts and spiking activity recorded from a sample electrode recorded in Manning et al's study. Panel A shows the normalized LFP power spectrum (black lines) and the mean broadband LFP power (red lines) for each of thirty consecutive 500-ms epochs. Panel B shows the neuron's spiking (black tick marks) and mean firing rate (blue lines) for these same epochs. Across these epochs, variations in broadband power were strongly correlated with simultaneous variations in the neuron's firing rate (Pearson's $r = 0.92$). Looking at the figure, one may observe that both broadband power and neuronal firing rate exhibited local maxima at 0.5 s, 4 s, 6.5 s, and 10 s, and both had local minima at 1.5 s and 5 s. Variations in LFP power were not limited to particular narrow frequency bands, but rather appeared as overall broadband shifts in the entire power spectrum (brown lines in panel A).

To determine whether this pattern was robust across the entire recording session, Manning and colleagues examined broadband LFP power and firing rate for each of the half-second epochs recorded for this neuron. Data from each epoch appear as a point in Figure 2.12a, where the horizontal coordinate indicates the firing rate and the vertical coordinate indicates the normalized broadband power. Across the entire recording session, these points were clustered along the diagonal, indicating that neuronal firing rate was positively correlated with LFP broadband power (Pearson's $r = 0.6$). Figure 2.12b depicts this relation in a different manner, showing the mean LFP power spectra for each of five groups of epochs where this neuron had different firing rates (different colors in panel a). As this neuron's firing rate increased, the LFP power spectrum exhibited a proportional upward shift at *all* frequencies.

Manning et al next sought to identify all neurons whose firing rates varied with broadband power (as in the example above) or with narrow-band power. Because broadband power is influenced by each narrow frequency band, disambiguating broadband and narrowband effects is critical for understanding the relation between neuronal spiking and LFP activity. To identify neurons exhibiting each of these patterns, they fit a bivariate linear regression model to the relation between firing rate and measures of both broadband and narrowband LFP power. For each neuron, they computed the firing rate for each half-second epoch and they also computed LFP power measured at the same electrode at five narrow frequency bands: delta (2–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–150 Hz), in addition to computing broadband power. For each neuron they then performed a set of bivariate regressions where broadband power and the mean power in one narrow frequency band were simultaneously used to predict the neuron's instantaneous firing rate. The two β coefficients estimated in each regression indicate the contributions of broadband activity and this particular narrowband frequency band towards each neuron's firing rate.

Combining the results of all five regressions for each neuron, Manning

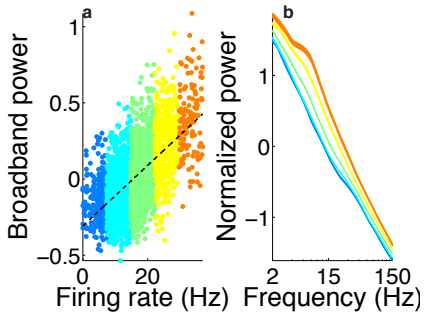

Figure 2.12: A representative neuron exhibiting a positive correlation between firing rate and broadband LFP power. a. Broadband power and firing rate for the neuron analyzed in the figure above. Each 500-ms epoch of the recording session is represented by one colored dot. The color of each dot represents its relative firing rate. Warm colors depict epochs with high firing rates, and cool colors indicate epochs with low firing rates. The dashed black line shows an ordinary least squares regression to these data. **b.** Average LFP power spectra for epochs with different firing rates. The same color scheme is used in both panels. As firing rate increases, the power spectrum exhibits a positive shift at all observed frequencies. The thickness of each line represents $\pm$ 1 SEM.

et al designated a neuron as a *broadband-shift neuron* when all five β coefficients for the broadband predictor were significantly different from zero in the same direction. They designated a neuron as a *narrowband-shift neuron* if, across the five regressions, either (a) one and only one narrow-band β coefficient was significantly different from zero or (b) exactly two narrow-band β coefficients at adjacent frequency bands (e.g. beta and gamma) were significantly different from zero in the same direction.

The broadband shift effect was remarkably unidirectional, with 92% of all broadband-shift neurons exhibiting this effect in a positive direction. In contrast, among narrowband-shift neurons, only 66% exhibited positive correlations. Figure 2.13 shows that broadband and gamma-band power were the two dominant LFP measures that positively correlated with firing rate. The proportions of neurons exhibiting these two effects were comparable to one another and were both significantly greater than the proportions of significant positive or negative correlations observed at other frequency bands

Assuming that brain function depends largely on the activity levels of neurons, the foregoing analysis provides valuable information on how to infer neuronal activity from changes in the local-field potential. Specifically, these analyses suggest that broadband increases in spectral power signal an increase in the firing rates of neurons in the region in which the LFP is being measured.

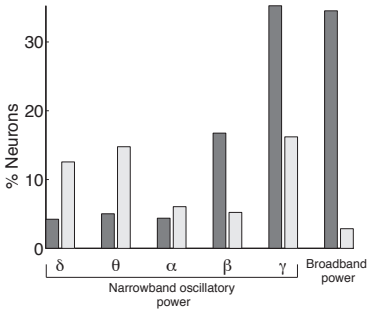

Figure 2.13: LFP components that predict firing rate. Dark gray bars indicate the percentage of neurons in each region that exhibited positive correlations between firing rate and a particular LFP feature; light gray bars show the percentage of neurons in each region that exhibited negative correlations. The bars on the left indicate the proportions of neurons whose firing rates were correlated with power in each narrow frequency band: delta (2–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–100 Hz). Each neuron may be counted in at most one direction (i.e., either positive or negative) per narrow frequency band. The bars on the right indicate the proportions of neurons whose firing rates were correlated with broadband power (i.e., broadband-shift neurons).

Gamma oscillations

Activity in the gamma band (typically defined in the 30-100 Hz range) reflects fluctuations in extracellular fields that are faster than those in the alpha band. Gamma activity typically increases during successful performance of a cognitive task and is, for example, enhanced during the learning of information that is later remembered relative to forgotten (Ezzyat et al., 2017). There is some debate about whether this type of increased gamma activity is truly oscillatory or instead reflects broadly asynchronous synaptic activity (Burke et al., 2015). Under the first scenario, synchronous gamma oscillations play a mechanistic role in cognitive processes such as memory encoding, in that the process depends on the coordinated firing of distinct neural networks (Jensen & Lisman, 2005); we discuss such a model of working memory encoding, the Lisman-Idiart-Jensen model, in more detail in the next section. Under the second scenario, increased activity in the gamma band reflects increased synaptic activity at the recorded location that is not synchronous. Under this view, increased gamma activity can be thought of as an index of the strength of underlying neural activity (Manning et al., 2009)–the implication is that gamma activation could reflect any number of underlying processes that are not memory specific. Instead, information about memory-specific processes is contained in the spatiotemporal pattern of gamma activity across the brain (see Chapter 3).

Theta-gamma interactions

In addition to the notion that oscillations at particular frequencies reflect (a)synchronous activity that *independently* support cognitive functions, there is also significant evidence for interactions between activity at different fre-

quencies. One of the most prominent examples is theta-gamma coupling (Colgin, 2015). Typically this refers to fluctuations in gamma power that consistently occur at a particular point in the theta cycle (theta phase-locked changes in gamma power). For example, in the human hippocampus, increased gamma power was shown to occur preferentially at the trough of the theta cycle during the encoding of remembered compared to forgotten words (Lega, Burke, Jacobs, & Kahana, 2015). Long-term potentiation (LTP), the primary mechanism for altering the connection strength between neurons, is dependent on the phase of theta, suggesting a reason for gamma activity to be coupled to the theta cycle in hippcampus (Hasselmo & Eichenbaum, 2005).

Electrophysiological models of memory

Oscillations play an important role in many neurophysiological models of human memory. By way of introduction to this literature, we will here review three major models that use frequency-specific neural activity to account for memory processes.

Lisman-Idiart-Jensen Model of Working Memory

Given the fact that neural signaling events such as action potentials occur on the order of tens or hundreds of milliseconds, a basic question about episodic memory that has concerned researchers is how these neural mechanisms can support the formation of associations between events that occur further apart in time. For example when listening to a professor's lecture, important information may be described over the course of tens of seconds or a minute. Although the information is spaced out in time, your brain is able to form a memory representation that associates all of the relevant information via strengthening of synaptic connections between neural assemblies representing each element of the memory. The fact that your brain can do this may not seem surprising, except that it is incompatible with long-term potentiation (LTP), the fundamental cellular mechanism that underlies hippocampal learning and memory (Markram, Lübke, Frotscher, & Sakmann, 1997), which serves to strengthen connections between cells that fire within 100 ms of each other.

In the Lisman-Idiart-Jensen (LIJ) model (Jensen & Lisman, 2005) proposes that individual elements of a sequential memory trace are maintained in a cross-frequency working memory buffer. This maintenance occurs in the cortical structures (such as entorhinal cortex) that provide input to the hippocampus, where LTP is induced. Once per θ cycle, the network of cells that represents each item is activated. This is proposed to occur at the gamma frequency, with later items being represented at later sequential γ cycles.

Because γ has a frequency of roughly 30 to 120 Hz, this model has several qualities that make it attractive as an explanation for how working memory processes support long-term memory formation. The γ period roughly corresponds to the inter-item separation time that has been identified in the Sternberg task (Jensen & Lisman, 1998). There are also roughly seven γ cycles for each θ cycle, which implies a physiological limit of about seven items on the number of items that can maintained in working memory, consistent with classic work on working memory capacity limitations (G. A. Miller, 1956).

Repeated activation of items at the γ frequency across multiple θ cycles also allows for items to be activated multiple times, increasing the strength of LTP.

Hasselmo Model of Hippocampal Encoding and Retrieval

Oscillatory synchronization has also been proposed to play a role in separating the processes involved in memory encoding and retrieval. This question is particularly relevant to theories of hippocampal function, since this structure is proposed to support both encoding and retrieval of episodic memories. However, the properties that would make the hippocampus a successful encoder of new memories, namely the ability to decorrelate similar inputs to create distinct memory representations, conflict with those necessary to retrieve memories given partial cues.

The Separate Phases of Encoding and Retrieval (SPEAR) model proposes that interactions between entorhinal cortex and hippocampus that support encoding and retrieval are instantiated at different phases of the theta rhythm. At the peak of theta, LTP is upregulated in the hippocampus, leading to association of presynaptic patterns in CA3 and postsynaptic patterns in CA1, both of which are driven by external input from entorhinal cortex. In contrast at the trough of theta, activity generated by previously stored memories in CA3 will drive the response of CA1 and its output to entorhinal cortex. Because LTP is minimized at the theta trough, the pattern evoked by CA3 activity is not encoded as a new memory in CA1 and the system is biased for retrieval (Hasselmo et al., 2002).

Complementary learning systems

One class of neurophysiological model of episodic memory concerns the process by which memories undergo stabilization as a result of consolidation. Standard systems consolidation theory (Alvarez & Squire, 1994) was initially motivated by observations of temporally-graded retrograde amnesia, wherein patients show a gradient of forgetting of previously learned memories with those encoded just prior to the the amnesia-provoking injury showing the most forgetting, and memories experienced further back in one's personal history showing relatively normal forgetting. Systems consolidation theory proposes that newly-formed memory representations are encoded as patterns of activity in the hippocampus which, over time, come to be represented in cortical networks. The process by which this occurs is through information transfer from hippocampus to cortex, hypothesized to happen during offline rest periods and during sleep. Computationally, this process is necessary because the hippocampus and cortex are specialized to learn at different rates (i.e. synaptic plasticity occurs at different timescales) with the hippocampus able to rapidly acquire memory representations within a single trial, while cortex requires multiple presentations or exposures to patterns of information in order to acquire stable representations (Norman & O'Reilly, 2003). Physiologically, high-frequency bursts of activity in the hippocampus (sharp-wave ripples) are thought to reflect communication between the hippocampus and cortex, with low-frequency oscillations (e.g. θ) proposed to coordinate activity between hippocampus and cortex (Buzsáki, 2015; Ji & Wilson, 2007; Siapas, Lubenov, & Wilson, 2005). Single-unit studies in ro-

dents also provide evidence for systems consolidation theory, demonstrating replay during sleep of place cell sequences that were active during pre-sleep experiences

A Big Data Affair

We have now seen that human electrophysiology is a big data affair: brain recordings generate vast data sets that we wish to relate to the stimulus and response events recorded during a memory experiment. Parts 2 and 3 of this book provide a hands-on introduction to the quantitative and statistical methods used by researchers when analyzing these data. To facilitate "active" learning of this material, we have chosen several large publicaly available datasets as the testbed for the methods surveyed in upcoming chapters. Below we briefly summarize each dataset, explaining its motivation and listing some key findings. We refer to appendices which contain the methodological details needed to carry out further analyses of these data. All datasets maybe downloaded through the UPenn Computational Memory Lab website `http://memory.psych.upenn.edu`.

Delayed Free Recall of unrelated and categorized word lists

. In this experiment, subjects contributed multiple sessions of delayed free recall of either unrelated or categorically organized word lists (Experiment Codes FR1 and CatFR1, repsectively). Subjects were neurosurgical patients undergoing intracranial electroencephalographic monitoring as part of clinical treatment for drug-resistant epilepsy. Data were collected as part of a multi-center project designed to assess the effects of electrical stimulation on memory-related brain function. [7] Electrophysiological data were collected from electrodes implanted subdurally on the cortical surface as well as deep within the brain parenchyma[8]. In each case, the clinical team determined the placement of the electrodes so as to best localize epileptogenic regions. Subdural contacts were arranged in both strip and grid configurations with an inter-contact spacing of 10 mm. Most subjects also had temporal lobe depth electrodes with 5 mm inter-contact spacing. Intracranial data were referenced to a common contact placed either intracranially, on the scalp or mastoid process. We provide both these raw data as well as data rereferenced using a bipolor montage (Burke et al., 2014).

Each subject participated in a delayed free-recall task (illustrated in Figure 3.14) in which they were instructed to study lists of words for a later memory test; no encoding task was used. Lists were composed of 12 words chosen at random and without replacement from a pool of high frequency nouns (either English or Spanish, depending on the participant's native language; `http://memory.psych.upenn.edu/WordPools`). Each word remained on the screen for 1600 ms, followed by a randomly jittered 750-1000 ms blank inter-stimulus interval (ISI).

Immediately following the final word in each list, participants performed a distractor task (20 seconds) consisting of a series of arithmetic problems of the form A+B+C=??, where A, B and C were randomly chosen integers ranging from 1-9. Following the distractor task participants were given 30 seconds to verbally recall as many words as possible from the list in any order; vocal responses were digitally recorded and later manually scored

[7] Data were collected at the following centers: Thomas Jefferson University Hospital (Philadelphia, PA), Mayo Clinic (Rochester, MN), Hospital of the University of Pennsylvania (Philadelphia, PA), Emory University Hospital (Atlanta, GA), University of Texas Southwestern Medical Center (Dallas, TX), Dartmouth-Hitchcock Medical Center (Lebanon, NH), Columbia University Medical Center (New York, NY), National Institutes of Health (Bethesda, MD), and University of Washington Medical Center (Seattle, WA). The research protocol was approved by the IRB at each hospital and informed consent was obtained from each participant.

[8] Anatomical localization of electrode placement was accomplished using a two step process. First, hippocampal subfields and MTL cortices were automatically labeled in a pre-implant 2 mm thick coronal T2-weighted MRI using the automatic segmentation of hippocampal subfields (ASHS) multi-atlas segmentation method (Yushkevich et al., 2015). Following this automatic labeling procedure, a post-implant CT scan was co-registered with the MRI using Advanced Normalization Tools (Avants, Epstein, Grossman, & Gee, 2008). Electrodes that are visible in the CT were then localized within the MTL subregions by a pair of neuroradiologists with expertise in MTL anatomy. The whole brain cortical surface was also obtained from a pre-implant volumetric T1-weighted MRI using Freesurfer (Fischl et al., 2004), and subdural electrodes were separately co-registered and localized to cortical regions using an energy minimization algorithm (Dykstra et al., 2012).

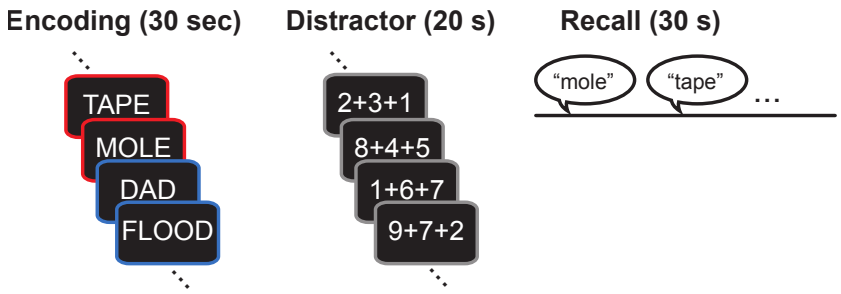

Figure 2.14: Illustration of the behavioral task in the RAM FR study. Participants are presented lists of 12 words one at time during encoding, followed by an arithmetic distractor (20 s), and finally the recall period (30 s).

for analysis. Each session consisted of 25 lists of this encoding-distractor-recall procedure. A subset of subjects completed additional sessions of the free recall task using categorized word lists, which were included in the electrophysiological analyses.

PEERS: The Penn Electrophysiology of Encoding and Retrieval Study

In this scalp EEG study of episodic memory, we sought to assemble a large dataset in which each subject contributed enough data (576 lists across 24 sessions) so that all statistical analyses of interest could be performed on a single subject. Initially, the plan was to recruit just a small sample of subjects (8 or 12) but after three years of testing, we assembled data on 76 subjects who each completed the full experiment. Details of the experiment can be found in prior publications (e.g. Kahana & Aggarwal, 2018), but we summarize the key experimental methods here.

Participants were recruited for the study through a two-stage process. First, right-handed native English speakers between the ages of 17 and 30 were recruited for a single session designed to introduce them to EEG recordings and the experimental task. Participants who completed this introductory session were invited to enroll in the full study, on the condition that they did not make an excess of eye movements during item presentation epochs of the experiment. Approximately half of the subjects recruited for the preliminary session agreed to participate in the multi-session study. In the multi-session study, participants completed a delayed free recall experiment consisting of 23 experimental sessions, each lasting between 1 and 2 hours.

Each of the 23 sessions began with a preliminary task in which participants saw a screen containing a fixation cross in the center, along with an asterisk above, below, and to the right and left of it. A series of auditory commands instructed the participant to make specific eye movements.

Following the preliminary task, participants completed a series of 24 free recall trials, with 24 words presented on each trial. Each trial began with a 10-second countdown, which was displayed on a computer screen in white text on a black background. Participants were permitted to pause and resume this countdown at any time by pressing a key. After the countdown was complete, a fixation cross appeared on the screen for 1500 ms. On half of the trials, the fixation cross was immediately followed by the presentation of the first word in that trial's list. On the remaining 12 trials, the fixation cross

was instead followed by a 24-second math distractor task. After this pre-list distractor task the screen went blank for a jittered 800-1200 ms (uniformly distributed), followed by the presentation of the first word. Trials were randomly assigned to one of these two conditions with the stipulation that the first trial should never include a pre-list distractor.

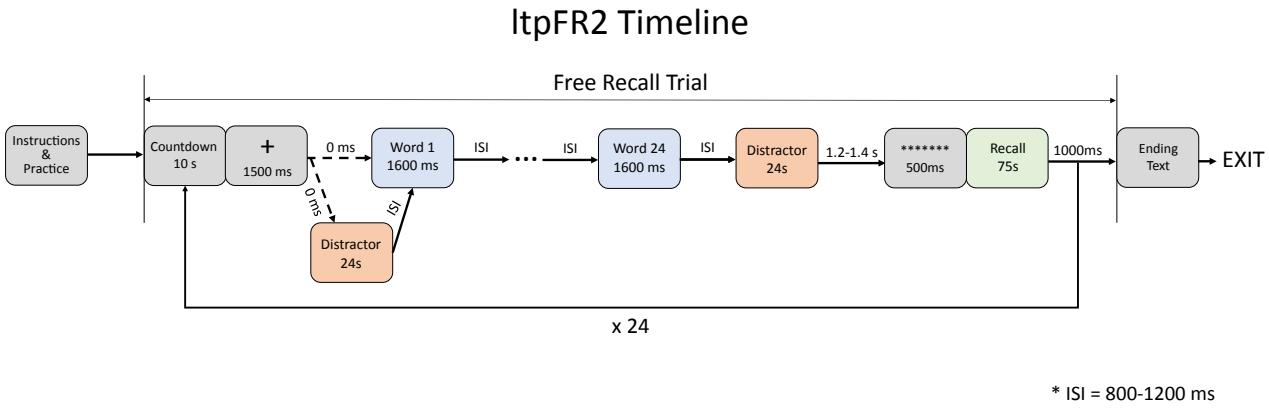

Figure 2.15: Timeline of the PEERS task. This timeline illustrates the main recall task in the PEERS experiment. Participants completed 24 trials, each of which contained 24 word presentations, a post-list math distractor, and a recall period. Half of the trials also contained a pre-list distractor.

All 24 words were presented onscreen, one at a time, in white text on a black background. Each word remained onscreen for 1600 ms and was followed by a jittered inter-stimulus interval of 800–1200 ms. After the inter-stimulus interval following the final word in the list, participants performed a math distractor task for 24 seconds. This post-list distractor was followed by a 1200–1400 ms delay, after which a tone sounded and a row of asterisks appeared onscreen for 500 ms, indicating the start of the free recall period. Participants were given 75 seconds to recall aloud as many of the words from the current trial as possible, in any order. A fixation cross was displayed onscreen for the duration of the recall period. Once 75 seconds had elapsed, the recall period ended and a blank screen was displayed for 1000 ms, marking the end of the trial. Following the 8th and 16th trials, participants were given a short break (approximately 5 minutes) during which experimenters entered the testing room to check on the EEG recording and adjust electrodes as necessary. The countdown for the next trial began once the experimenters had left the room. On all other trials, this 1000 ms blank screen was immediately followed by the countdown for the next trial.

Both the pre-list and post-list distractor tasks consisted of answering math problems of the form $A + B + C\ =?$, where A, B, and C were randomly generated positive, single-digit integers. Math problems were displayed onscreen, one at a time, in white text on a black background. Participants were instructed to type the answer to each equation as quickly and accurately as possible. Each time the participant entered an answer, the current problem was removed and replaced by a new problem. New problems would continue to appear until the full 24 seconds had elapsed, at which point the final problem was immediately removed from the screen.

Each session required $24 \times 24 = 576$ words. A single pool of 576 words was used for all 23 sessions. That is, participants saw the same 576 words in each of their sessions, but the grouping of these words into lists was

randomized for each session. Semantic relatedness between all words in the word pool was determined using the Word Association Space (WAS) model described by (Steyvers, Shiffrin, & Nelson, 2004). WAS similarity values were used to group words into four similarity bins (high similarity: $\cos\theta$ between words > 0.7; medium–high similarity: $0.4 < \cos\theta < 0.7$; medium-low similarity: $0.14 < \cos\theta < 0.4$; low similarity: $\cos\theta < 0.14$). Two pairs of words from each of the four bins were arranged such that the words of one pair occurred at adjacent serial positions and the words of the other pair were separated by at least two other items. These eight word pairs comprised 16 of the 24 words in each list. The remaining eight words in each list were selected and positioned at random.

Subjects were tested using one of two recording systems: the Electrical Geodesics HydroCel Geodesic Sensor Net (HCGSN) system and the BioSemi ActiveTwo system. Subjects with IDs between LTP093 and LTP330 were tested using the EGI system; subjects with IDs LTP331 and higher were tested using the BioSemi system.

[9] EGI system specifications

[9] EEG measurements were recorded using Geodesic Sensor Nets (GSN; Netstation 4.3 acquisition environment, from Electrical Geodesics, Inc.). The GSN provided 129 standardized electrode placements across participants. All channels were digitized at a sampling rate of 500 Hz, and the signal from the caps was amplified via either the Net Amps 200 or 300 amplifier. Recordings were referenced to Cz (electrode 129 on the GSN), and later re-referenced to the average of all electrodes To help identify eyeblink and other movement artifacts, electrooculogram (EOG) activity was monitored bipolarly using right and left electrode pairs on the GSN (electrodes 8 and 126 on the right; electrodes 25 and 127 on the left).

Biosemi system specifications EEG measurements were recorded using BioSemi ActiveTwo systems alongside BioSemi's ActiView recording software. Recordings were made from 128 standardized electrode placements across participants. The signal was amplified and digitized by an ActiveTwo AD-box at a sampling rate of 2048 Hz. The ActiveTwo system uses a feedback loop of two electrodes in place of an ordinary ground electrode (Luck, 2014, Box 5.3): a passive *driven right leg* (DRL) electrode injects a small current to move the the potential of the subject towards that of the amplifier circut and an active *common mode sense* (CMS) electrode serves the function of an ordinary ground electrode (the potentials between all other electrodes and the CMS electrode are recorded). Referencing takes place offline (we referenced to the average of all electrodes). To help identify eyeblink and other movement artifacts, electrooculogram (EOG) activity was monitored using additional electrode pairs at each eye.