

1.1 Foundations

Walle! walle,
Manche Strecke!
Daß, zum Zwecke,
Wasser fließe
Und mit reichem, vollem Schwalle
Zu dem bade sich ergieße.

Stehe! stehe!
Denn wir haben
Deiner Gaben
Vollgemessen! -
Ach, ich merk es! Wehe! wehe!
Hab ich doch das Wort vergessen!

From *Der Zauberlehrling*, J. W. von Goethe.

(Sweep! sweep // do not stop, // To reach my goal // let water flow // so streams and
torrents // fill the bath.
Stop! stop! // for we have // your gifts // in plenty! - // Oh, I see! Woe! woe! // I
no longer know the word!) Author's translation.

When the sorcerer's apprentice lost control over the demonic broom, the situation he had himself created continued regardless of his changing desires. Something similar happens when we try to change the way we respond to our environment. The existing state of mind seems to resist being overwritten. Once we start sweeping, there is a tendency to keep at it, even if we would really prefer to stop and do something else. But eventually the sorcerer is called back to use the magic word and stop the task to which we had set the brooms of our mind.

The main question this thesis aims to explore is what the magic word is. What allows us to change the way we react to the world? The studies described in the Experiments section are explorative attempts to provide information on this and related questions. In these experiments, analyses were done of the oscillatory behavior of the electric potentials generated by brain cells while subjects perform a variety of tasks. The reasons why such data might help understand how we change our mind are provided in the introduction. First, the concepts of information, control and encoding will be explained. These concepts are used in many if not all of the cited studies, but also provide an unambiguous way to talk and think about complex behavioral processes. Second, the main experimental paradigm of the experiments will be introduced: task switching. In this section, behavioral studies will be discussed. The following section is concerned with the organization of the brain and how that relates to task switching. Studies of brain activity look at the processes between the external events of stimuli and responses, and so could help understand behavioral results concerning relations between stimuli, responses and experimental manipulations. In following sections results on oscillatory neural behavior are discussed, organized by the three frequency bands of interest in the experiments. These sections, taken together with the foundational concepts, behavioral results and principles of brain organization, provide a framework from which tentative hypotheses arise and within which empirical data can be interpreted.

1.1.1 Information

Intuitively, information is a measure of how much a message tells you about something. For instance, if I tell you that a loaf of bread costs between two and three euros, I tell you less than if I tell you that it costs precisely 1.35 euros. If every time Pinochio lies, his nose grows by x cm, then the length of his nose tells you something about how many lies he's told. If Pinocchio tells you a loaf of bread costs 1.35 euros, you know less if his nose grows at that point than if it doesn't. A similarly intuitive idea of information processing is that that is what happens when you use what a message tells you. For instance, if you know the price of bread, you would make sure to bring at least that amount of money to the shop to buy a loaf. Information can also be communicated: messages can be passed along and maybe change their form. For instance, I could receive a letter by post, tell my neighbour what it says, and he could tell someone else via e-mail. Such ideas about information are relevant to the kind of self-control described above. The information conveyed by stimuli must be processed by the brain to eventually be used to choose an appropriate response. How the information is processed determines what response we choose. So, when we try to change the way we respond to events, we are actually trying to change the way we process information. But in the face of the complexity of behavioral patterns and neural organization, such intuitive ideas seem too vague to be of much use. Almost anything could be described as "information processing" - so by itself the term says very little.

A more rigorous theory of information was developed by Shannon [209] and Wiener [240]. This theory defined information in such a way that it became a formally defined entity and a precisely computable quantity. This allowed, for instance, the informativeness of two kinds of messages to be compared, even if their generation was complex and the differences subtle. It also allows the reduction of many kinds of mechanisms, capabilities, requirements, changes over time and patterns of activity to their characterization in terms of information processing, with sufficient detail to distinguish between and define specific cases. For instance, the brain contains on the order of 10^9 neurons. To demand each individual neuron and synapse to be described in all its uniqueness makes no sense and would bring neuroscience to a grinding halt. To describe each neuron too simplistically - as a point in space for example - would provide no understanding of how they subserve what the brain does. Considering the neuron as a precisely definable type of information processing unit is one of the steps that allows work towards an understanding of the principles of brain behavior. Information theory is in a sense a new language in which we can talk about patterns for which natural languages have insufficient words and rules of grammar. Because of its rigor, a statement made in information theoretical terms has immediate and unavoidable consequences that can be tested. Because of the power of the mathematical language, intuition-surpassing extrapolations and inferences can be made - for instance from one neuronal unit to many.

So at least a rudimentary understanding of information seems desirable, if not necessary, to try to understand what the brain does. Fortunately, the basic theory is simple to understand, needing only three building blocks: probability, surprise and entropy. Probability is a measure of how often a certain event occurs, expressed as a number between 0 and 1. For instance, when adding the values of two six-sided D6 dice, the event "the sum is smaller than 4" is fulfilled in three of the 36 possible outcomes. So, if all the outcomes occur equally often, the probability is $\frac{3}{36} = \frac{1}{12}$. The measure of surprise is directly mapped to a probability p by the function $-\log p$. So, when an event has a probability of 1, the surprise when it happens is 0. The closer the probability of an event gets to zero, the larger the surprise when it does in fact happen. The entropy, symbolized by S , is the weighted sum of surprises over a set of events: $S = \sum_{\text{event}} p(\text{event}) * \text{surprise}(\text{event})$.

This provides the expected value of surprise when an event occurs, that is, the average surprise over a large number of events drawn from the set. For all exhaustive and distinct sets of events (of which one must and only one may occur at a time), the summed probability is 1. Entropy, in contrast, provides a distinction between such event sets. The surprise, or uncertainty, involved in tossing a coin works out as $\frac{1}{2} + \frac{1}{2} = 1$ bit, bits being the unit of entropy. The entropy of rolling a four-sided die is 2 bits. For an eight-sided die the entropy is 3 bits, which shows the meaning of the bit measure: an increase in entropy by one bit is equivalent to doubling the number of alternatives of the set.

The formal definition of information is: the reduction of entropy due to a message. Once you see that the tossed coin is heads, the probability distribution becomes "p(heads) = 1", with an entropy of zero, so that 1 bit of entropy is lost. Rolling the eight-sided die reduces uncertainty by 3 bits, and so forth. For messages that are not 100 % reliable, the conveyed information is reduced by the remaining entropy following a response, averaged over responses.

1.1.2 Coding

Coding is the form of the message that is used to convey information. For instance, consider the intensity of light falling on a receptive cell. Even if it is known that the cell transmits the information, it could do so in different ways. Its firing rate could increase, by some monotonic function, with intensity. It could fire sooner after stimulation (this is an example of a time, or temporal code). It could fire a burst of increasing length or frequency. If the cell connects to a group of cells, the proportion of cells in the group that start firing could also encode the intensity. The firing rate of specific cells is a very well-studied code in the brain (the in vivo studies of section 1.3.2 all look at this code). However, so-called population [29] and temporal [235] codes are also used in the brain.

First, note that distributed and population coding are not synonymous. A distributed code involves multiple messengers, or signals. The abstract event "7, 3" could be encoded into a single value by a coding scheme of the form "message = (10 * number 1) + number 2", combining the information into a single message. It could also be encoded into two messages, one for the first and one for the second number. As illustrated in this example, while distributed coding requires more messages, the complexity of the encoding per message can be kept simpler. The neurons in our central nervous system are of course the instance of distributed coding of most interest to this text. An example of distributed coding is the representation of whisker deflections in rat somatosensory cortex [179] [180]. Each whisker is mapped, via the thalamus (see section 1.5.4), to a specific cellular unit (or cortical column [160], see 1.3.2) but also evokes activity in neighbouring units. The response to whisker deflection is initially restricted to the central unit but spreads to neighbouring units during the 10 ms following stimulation. Which stimulus has been presented can be reconstructed from the messages of the neighbouring units alone, indicating a spread of information over the population [179]. Extra information is provided by the timing of the first post-stimulus spikes [180]: the more central the cell, the quicker the first spike. The transmission of information in this way - by spike timing - is an example of a temporal code. Another form of temporal coding, synchrony, has been shown to play a distinct role from rate coding. Synchrony is a form of temporal coding in which neural events provide the reference time, as opposed to external stimuli. In motor cortex, the modulation of rate and spike synchronization have been shown to be differently related to motor functions [195]. Monkeys were trained to perform a delayed-response hand movement task, containing a preparatory signal and a response signal, presented after a

delay of 600, 900, 1200 or 1500 ms. Each delay was equally probable. Neurons in primary motor cortex showed an increase in synchronization, independently from changes in firing rate, around the delay times, even when no response signal was given. Rate modulation was found around actual responses. Thus, spike synchronization was related to internal, behaviorally relevant events (i.e. expectation and preparation), while rate modulation was related to producing a movement.

A population of neurons that responds to a stimulus by a distributed code does not, by itself, imply the use of a population code. A population code requires synergy [29]: the population response must convey more bits of information than the summed non-redundant information of the separate signals. This happens when the relationship between spikes of different neurons defines the events carrying information. For instance, in a population of four neurons, cells 1 and 3 could fire in response to stimulus A, and cells 1 and 4 to stimulus B, and so forth. "Reading" cell 1 could then not distinguish between stimuli A and B, while reading the combined response could unambiguously distinguish stimuli. In the rat somatosensory data, the late spikes in the neighbouring units could be due to any whisker with a neighbouring central unit. It is their pattern - around a given central, fast-spiking unit - that uniquely determines to what whisker they are responding, although in this case, the late-spiking pattern supplied no more information than the fast, central spike. In the antennal lobe of locust olfactory cortex, the response to odors consists of stimulus-dependent, spatially distributed spiking patterns [127]. Whether this is due to information encoded independently in spiking patterns over time for each cell individually, or whether "reading" the whole pattern is necessary to reconstruct stimulus information, was not tested directly. The data do, however, at least illustrate the feasibility of population codes.

Population coding leads to the superposition problem, which can be overcome using temporal coding, in particular synchronization [212]. For instance, stimuli in the environment do not occur one at a time or in isolation, and so their presence must be encoded simultaneously. How are the distributed signals that encode the component parts of different stimuli separated from each other? The superposition problem can be solved by combining population coding with an extra coding dimension for encoding set membership; this combination has been termed assembly coding [212] [196]. The precise timing of neurons' spikes may be used to define sets, as neurons are sensitive to the precise (millisecond scale) timing of their inputs. This sensitivity is due to coincidence detection [10]: the peak of the superimposed post-synaptic potentials is higher for synaptic events that occur closer in time. A set-coding dimension that uses spike timing would separate a population of active neurons into subsets, the elements of which would fire with a specific phase difference relative to each other. If that phase difference is zero, then the set coding dimension is synchrony. Synchrony has been argued to be well-suited for assembly coding: set membership is communicated quickly - one synchronous burst would be sufficient - and the number of overlapping assemblies is limited only by the time needed to separate the post-synaptic effects of such bursts [212]. Assembly coding using synchrony has been shown to arise in visual cortex. In monkey visual cortex, two sets of neurons were selected with different preferred orientation of light bars, but with overlapping receptive fields [123]. When a single bar with an orientation to which both sets of neurons responded was presented, all measured neurons fired synchronously. When two light bars were superimposed, one at each preferred orientation, the neurons in the two groups fired synchronously with each other, but lost their synchrony with the other group. In cat visual cortex [56], a similar study has been performed using either a single vertical bar moving across the screen, or two smaller vertical bars moving in opposite directions. Neurons with receptive fields through which the bars passed showed synchronous activity for the single bar, but unrelated spike timing for the separate bars.

These findings agree with a phase-coding solution to the superposition problem.

One form of temporal coding is phase coding. In this case, coding time is defined relative to some reference oscillation. Phase coding has been used to establish relations between representations other than only set membership. In a study on proposition representation (in an abstract system not intended to be very similar to the brain), statements such as "John loves Mary" and "Mary loves John" were separated from each other by the timing of the elements' activation [94]. During cycles similar to musical measures, terms such as "beloved" and "lover" were activated at different times, and names such as "John" and "Mary" were active during the active phase of one or the other term. In the locust olfactory system, cells in the mushroom body, to which neurons of the antennal lobe project, fire in a specific phase region relative to a 20 Hz oscillation of the local field potential [178]. Only input to the mushroom body that is correctly timed to this window of opportunity can contribute to further spikes. Thus, while a complex spatiotemporal pattern of spikes encodes the odor, the phase of spikes encodes the salience of different parts of that pattern.

All in all, population and temporal coding have the potential to be an important part of information processing in the brain. Evidence exists, at least for primary sensory cortices, for a richer repertoire of coding mechanisms than only rate coding. Later sections will review studies of distributed and temporal coding at larger scales than the *in vivo* studies above.

1.1.3 Information and working memory

Many studies cited in this thesis refer to working memory, which has been defined as "the temporary storage of information that is being processed in any of a range of cognitive tasks" [11]. In this definition, working memory is an act - "storage" as opposed to "storage space". The existence of such acts can be inferred from the performance of tasks with suitable demands, e.g. "remember the following random list: ...". However, working memory has also been defined as a system, comprising a "central executive", "visuospatial sketch pad" and "phonological loop", each of which assigned functions involved with, respectively, attentional control, the temporary storage and manipulation of visual and spatial information and the maintenance of items in subvocal speech [12]. So it is not clear whether "working memory" should be conceived of as the act of temporary storage, the place where information is temporarily stored, or the systems operating on such information. Many variations of the definition also exist, e.g. "the ability to transiently hold and manipulate goal-related information to guide forthcoming actions" [54].

The reason working memory was postulated was as an elaboration of a unitary short-term memory, which could not explain dissociations in neuropsychological findings within the set of short term memory dysfunctions [12]. It seems likely that working memory defined as an ensemble of systems would have to be upgraded to include every modality introduced by new tasks (e.g. comparing two smells, determining whether a sequence of touches is a sensory palindrome, and so forth). The underlying idea of working memory concepts appears to be the existence of distributed short term memories combined with at least one and possibly a unitary organizing structure. A different kind of definition of working memory can be given in terms of task demands, in contrast with the inferred structures and processes involved in performing those tasks. Working memory tasks, such as those listed above, involve a certain type of demand: the processing of temporarily stored information [11]. In this text, unless otherwise stated, "working memory" refers to the task demands of temporary information storage and the manipulation of such information.

What does it mean to say that a task demands the temporary storage of information? Take the following two situations. First, let two events A and B occur at time t_0 . Then information

of the type "A == B" might need to be encoded into a response. Next, let an event A occur at time t_0 and be encoded in some form at place p_0 . What happens if at a later time t_1 a new event B occurs and the task still demands the encoding of the event A == B? The information of event A occurring must be communicated over time to t_1 , when it can be processed together with event B to produce the desired response. So, storage of information will be defined here as communication over time. Temporary storage occurs when stored information is lost after some duration or processing step. Note that this actually encodes a specific kind of information, namely timing: if the information is still active, it must be due to a recent event. The complementing kind of storage, long-term storage, encodes what has happened at some point, or over some period, in the past. Prefrontal cortex has been shown to play an important role in communication over time and in goal-directed changes in how information is processed, as discussed in section 1.3.2.

1.1.4 Neural channels for communication over time

The temporary storage and manipulation of information requires a certain kind of encoding. After a stimulus is encoded into a neural pattern, that pattern must communicate itself to a future point in time. At that point, the information can be processed further, e.g. when more information becomes available. The typical example of such temporal communication is when a cue is given that must be combined with a subsequent stimulus, presented after a delay, to encode the event "the cue was (or consequent context is) X and the stimulus is Y". Similarly, since manipulation of information takes time, an enduring encoding is a prerequisite for that second aspect of working memory. What form could such encoding take? Two mechanisms which have been used in neurocomputational models of working memory [54] are recurrent excitation and cellular bistability. The ideas involved in these mechanisms may help understanding task switching and its underlying brain activity.

Recurrent excitation involves a pattern of activity feeding back onto itself in such a way that it converges to a stable state over time (or iterations) [90] [83]. Note that communicating information via a distributed pattern of activity is a form of population coding. Other terms for recurrent excitation are reverberation [6], autoassociation [83] and attraction [6]. Hopfield networks [90] are a well-known example of a system of recurrent excitation [54] [6], which illustrate three important concepts: convergence, stability and temporal tuning. Hopfield networks consist of interconnected elements, termed neurons, which compute whether their weighted input exceeds a threshold. There is no direction in the structure of a Hopfield network, as in a neural network with an input and output layer connected either directly or via hidden layers; only time provides direction, in the way described below. Each element has a state variable, its activation, which is either one or zero. The connection, termed synapse, from a neuron i to a neuron j specifies the weight neuron j applies to neuron i 's activation value when it determines its input. The biological terminology was based on an analogy with the firing rate of a neuron as a function of the effect of synaptic events on its membrane potential. The elements are allowed to independently and stochastically determine at which times they summate their input and perhaps change their activation. The network encodes information through its convergence to one of a number of patterns, stored in the synaptic weights. The storage algorithm bases the weight between two neurons on the co-occurrence of their activations over a set of activation patterns to be stored. This choice of weights can be learned by the network through Hebbian learning, and once instantiated, leads to convergent network behavior. Convergence means that changes in the activation pattern remain low after a sufficient period of time following a given starting pattern, during which the eventual non-changing, or stable, pattern is reached. The converged pattern is said to have low energy, the energy of the

network at a certain time point being defined as a measure of how much the activation pattern will change in the next iteration, or over the following time period. Low energy therefore means that the activation pattern remains the same. The stored patterns on which the synaptic weights are based become stable patterns. If a pattern of activation is imposed on the network, it only has to be sufficiently similar to one of the stored patterns for the network to converge to the full stored pattern. This kind of pattern retrieval is termed content-addressable memory.

Note that in Hopfield networks the only direction is time, not any kind of flow from input to output layers within the network. As described in 1.3.1, this kind of "inwards-looking" processing may be an important part of coordinative brain function. Further, the abstract, computational neurons in the network could be hypothesized to represent meaningful functional units of the brain, implementing e.g. stimulus processing or response generation. Pattern completion in the network would in that case translate to stimulus - response mapping, and since various patterns can be stored, such a mapping of stimuli to responses would be flexible. The question of how task switching works in computational terms might then be related to the question of whether and if so how pattern completion involving sensory and motor regions is performed by the brain. If pattern completion is the underlying process involved when subjects prepare to perform an upcoming task, certain questions arise. First, something must determine a starting point for the network, which determines its nearest stable state. Note that the tendency to converge provides control as well as information storage and manipulation. Second, even if the end-state is determined, a period of time is required for convergence to take place. Third, if the same network is used for subsequent patterns, interference may occur, and the energy of the network must be increased sufficiently to break out of established stable states. Hopfield believed that the convergence behaviour in his networks would generalize to other specific implementations, and in line with that belief these questions about pattern-recognition, although inspired by the general behaviour of Hopfield networks, do not rely on their details.

Neuronal delay activity measured from cortex using single-unit recordings has been modelled using an attractor model [6]. In the experiment [155], visual stimuli were presented in a fixed sequence during training. Unexpectedly, delay activity for different stimuli was correlated with how close they were to each other in the learning sequence. The temporal closeness had been encoded into spatial similarity. An attractor model was built using excitatory integrate-and-fire neurons. The derivative of a neuron's current was a function of the weighted sum of the firing rates of the other neurons and a decay term. Patterns were presented as an additional term in the equation for the current derivative. A nonlinear current-to-rate transduction function was used. Unstructured inhibitory feedback was provided that led to an overall hyperpolarization driven by the mean activity in the pool of excitatory neurons. The synaptic matrix was learned using a periodic sequence of patterns. Weights w_{ij} between neurons i and j were 0, a or 1. If a pattern exists in which both neuron i and j are active, then $w_{ij} = 1$. If a consecutive pattern exists in which neuron i and j are active, one in pattern n and the other in pattern $n + 1$, then $w_{ij} = a$. Otherwise, the weight is zero. This kind of learning algorithm allowed the network to learn sequences and exhibit the same behavior as that found in the empirical data. When pattern n is presented and then removed, the network stays active in that pattern, but also activates the surrounding patterns, which in turn activate their neighbours. The learning parameter a determines the breadth of temporal tuning, i.e. the tendency of pattern n to evoke patterns that are nearby in terms of sequence position, or more generally, time. Note that, the stronger neighbouring patterns become activated, relative to the relevant pattern, the less specific the networks response is to the

actual stimulus, so that a trade-off appears to exist between the precision of representation and breadth of temporal tuning. This trade-off may play a role in the organization of prefrontal cortex (section 1.3.2). Another similarity between this model and prefrontal cortex is the ability of cells to code retrospectively and prospectively, that is, bidirectionally in time, just as was achieved by a non-zero temporal tuning parameter.

A problem with the specific Hopfield network as a model for working memory is that the synaptic weights must be learned for the network to function, while novel information should be able to be encoded in working memory with no or a minimal learning phase. If Hopfield networks are to play a role in working memory, it seems that either a) a large number of patterns must already be stored, or some further mechanism exists which either b) allows fast, transient changes in synaptic weights within the Hopfield network or c) flexibly connects the neurons in a central Hopfield network to peripheral, content-providing neurons (e.g. for perception and action), changing the meaning of the central patterns. However, no task is performed without either training (especially in animal studies) or an available analogue in long-term memory (especially in human studies), which may be important for this problem.

In cellular bistability models, oscillation due to the dynamics of realistic model neurons combined with recurrent inhibition is the key to memory maintenance [237]. As will be shown later (in 1.4, 1.5 and 1.6), neural oscillations are extensively involved in cognition, so such oscillatory mechanisms are of special interest. Recurrent inhibition provides a robust mechanism for the generation of oscillations [237]. The basic recurrent inhibition model reciprocally connects neurons belonging to two groups: one excitatory and one inhibitory. Excitation of the excitatory group results in spikes that cause the inhibitory neurons to fire. They in turn inhibit the excitatory pool, thus depriving themselves of input. Now the excitatory pool can fire again, and the oscillation enters its second period. The continuation of the oscillation requires either a continuous depolarization of the excitatory neurons, or long-lasting depolarizing current, such as caused by NMDA channels [237]. In the latter case, when spiking occurs in an interconnected group of excitatory neurons, both a post-synaptic depolarization within the excitatory group and recurrent inhibition are initiated. The depolarization may last long enough to still be able to reach spiking threshold after the inhibitory phase is over, which starts the cycle again. Recurrent inhibition causes synchronization because if neurons fire at all, that must occur in the time window between two inhibitory phases.

The presence of oscillations provides the potential for phase coding, as the effect of incoming excitations may be dependent on the phase of the oscillation, as shown in olfactory cortex in the locust [178]. The phase coding in locust olfactory systems has further been shown to be generated by recurrent GABA inhibition from lateral horn interneurons [178]. Recurrent inhibition has been argued to be part of the cause of the encoding of tones by single spikes in auditory cortex [51]. A further example of the importance of the timing of synaptic input relative to rhythmic inhibition will be shown in section 1.5.4 on thalamocortical behavior.

Cellular bistability provides a short-term memory mechanism: a pattern of activity can remain after a stimulus is no longer present, encoding the past event, and the memory lasts only as long as the oscillation, with no changes in synaptic weights. Oscillatory activity in post-synaptic potentials has also been associated with assembly coding, as a method to make patterns of synchrony within a network robust by coupling the precise timing of discharges to the oscillation and not to the perhaps noisy timing of individual synaptic events [212] [196]. Cellular bistability could play a further role in maintaining an assembly code over time. If bistable elements are part of a Hopfield network, pattern completion and temporal tuning could be added to a network of such elements'

capabilities. The channels provided by attractors and cellular bistability, together with assembly coding, provide a computational framework for thinking about cognitive control in reductionist terms, and will recur in the discussion of various fields of research.

1.1.5 Control and cognitive control

Similarly to information, control has an intuitive and a rigorous meaning. Intuitively, control is the ability to either change or stabilize something in accordance with our desires. A more rigorous development was initiated by Wiener in his book on cybernetics [240]. Wiener identified control with negative feedback applied to error signals. This idea contains four important elements. First, an error signal implies both information on some variable and a target or goal, deviations from which define the error. So control, as intuited, involves goals. Second, the negative feedback means that during control, such actions are taken that result in a decrease in error. Again, this formalizes an intuition: control should not let errors persist, and certainly not increase them - at least not as the final result. Third, an intimate relation with information is implied. The error signal must be communicated to the control system, but there is also a deeper identification. Control can also be described as the encoding of error signals into reactions, via such a coding that negative feedback occurs. In this way, control mechanisms implement this special kind of information processing.

The fourth consequence of defining control as negative feedback is the least obvious. When there is no error signal, no action is required; the situation could be described as satisfaction. The idea of dissatisfaction, or unpleasure, leading to action, is the basis of Freud's distinction between the primary and the secondary process [60]. The most primitive form of cognition, according to Freud, consists of the generation of random actions in response to unpleasant stimulation. The action leading to removal of the stimulus would become associated in memory with both the stimulus and the pleasure (e.g. of reducing hunger or pain). This primitive cognition could then fall into the trap of achieving a hallucinatory pleasure via such associations between thoughts instead of by responding adaptively to the environment. The primary process of the mind is this activation of ideas solely on the basis of associatively reaching pleasurable memories. The secondary process was suggested to have evolved to prevent primary process thinking from leading the organism into hallucination. In other words, it had to prevent pleasure from being achieved in a way that did not take reality into account. To do this, new rules for associating ideas had to be imposed. While the primary process simply spreads activation along lines of association until pleasure-nodes are reached, the secondary process must block those lines that would work less well in reality than in fantasy. So, at the basis and the pinnacle of thought, lies control. Primary process thinking starts by responding to an error signal; secondary process thinking controls hallucinatory pleasure seeking via the inhibition or biasing of primary process activity.

It is interesting to note that these ideas were published in 1900, in "The Interpretation of Dreams" - that is, before information theory, control theory, behaviorism (especially operant conditioning) and computation in neural networks, and in the same period as the establishment of the neuron theory, the cellular theory of the nervous system. Yet such important ideas were not only foreshadowed but combined, in such a way as to also link simple underlying processes to motivation and emergent patterns of thought (e.g. condensation). While this text will not delve further into the original psychodynamics, it should be noted that well-known cognitive dual-process hypotheses (discussed next) and an organizing function of prefrontal cortex (section 1.3.2) on which the experiments in this thesis are based, seem to lie close to the primary and secondary process described above.

Dual-process hypotheses emphasize differences between two ways in which people can process information. In Schneider and Shiffrin's two-process theory of human information processing, controlled processing is defined against the background of a long-term memory consisting of broadly defined units or nodes [204] [205]. These units are abstract representations of mental events, including perceptions and actions but also complex mental events, such as action sequences or shifts of attention. Information processing is defined, again at a high level of abstraction, as the activation of a sequence of such units. The claim of the theory, initially based on visual search experiments using a Sternberg task, is that sequences of nodes can be activated in two ways. Sequences can be initiated and completed regardless of current goals, if they are sufficiently well-learned. Such activations are termed automatic processing. Automatic processing describes the pattern of behavior when subjects consistently search for the same set of stimuli in a display: under such consistent mapping conditions, detection times are independent of the memory and probe set. Controlled processing is the activation of a sequence that is temporarily set up to perform a task; this does not require the sequence to be part of long term memory itself, but does require attention. Under varied-mapping conditions, when subjects had different memory sets on each trial and so could not automatize processing, increased probe and set size caused longer response latencies. Note that the metaphor of activation sequences in connected nodes are a geometric abstraction of successive instances of encoding, or chains of communication, and not a direct description of spreading activation in a neural network. In visual detection, Posner et al. contrasted orienting and detection in an analogous way to controlled and automatic search [190]. Orienting was defined as both overt behavior bringing sensory and motor systems into an advantageous position (e.g. moving the eyes towards objects), as well as inferred covert preparation. Detection, in contrast, concerned the arrival of sensory input into an information processing system that could output arbitrary (e.g. experimenter-determined) responses. Orienting, based on pre-stimulus cues, was shown to occur before detection. This led to the suggestion that habitual responses, such as orientation to locations, can occur independently from the nonhabitual responses involved in detection. Thus, again, sequences of mental events were proposed to occur either because they are well-learned, or because they are specified by a current, flexible goal concerning stimulus - response mappings. The same basic distinction, although stated in terms of systems instead of processes, is made in Norman and Shallice's model of schemata and the supervisory attentional system [167]. Schemata are abstractions based on experience, including habitual behavioral patterns in response to stimuli. Since people do not always respond in their most habitual manner - for instance when asked to perform a Stroop task - competition between schemata, termed contention scheduling, does not seem to provide a full explanation for behavior. A supervisory attentional system was proposed to exist, and to be able to bias contention scheduling when such competition was inadequate to achieve desired performance. Again, it could be said that an important difference between the two systems is the degree to which current goals concerning external behavior govern internal processes.

Controlled and automatic processing have been implemented in a hybrid symbolic - connectionist architecture called CAP2 [205]. This architecture consisted of a modular data matrix and a control system. Automatic processing is the basic information transmission by the autoassociative, connectionist modules, based on the coding of the priority of their input. Controlled processing occurs when a module that would otherwise not transmit information is provided with an output gain signal by the control system. The model involves various subdivisions and signals, but a perhaps important point about its structure is the separation of the data matrix and control system. Note that controlled *processing* does not happen in the control *system*: controlled processing hap-

pens in the data matrix, modulated by the control system. Such a separation of organization and representation will resurface in the discussion of the function of prefrontal cortex (section 1.3.2).

So, in the dual-process models, information has a default, robust way of being processed, based (in these models, but see below) only on learned associations, but the way information is processed can be changed, using temporary, goal-directed associations. In terms of control, the error signal that arises during controlled processing is the deviation of the desired activation sequence and the actually activated pathway. Since the long-term associations do not disappear when undesired, control must be exerted to inhibit escapes from the goal-related sequence. This kind of control is termed cognitive control. In neural terms, cognitive control is the control of the way information is or will be processed in the brain, where error is caused by some form of automatic processing. Task switching is an instance of cognitive control that occurs around the time the target is changed, and so a previously satisfactory state becomes encoded as an error. So, however the control is implemented, it must lead to such negative feedback that a previously stable state is replaced by a new state of how information will be processed. The processes of overcoming stability and reconfiguring the chains of communication, and how these processes occur in response to changing goals, are the themes of task switching research, as discussed below. In task switching, automatic tendencies or biases are not due to habits or well-learned memories, but to transient states due to recent task performance. What remains of the controlled - automatic distinction is not so much the habitual - arbitrary dichotomy, but whether activity is or is not a consequence of goals at the level of stimulus - response mappings.