1 Introduction

The books, papers, and lectures which I appreciate most start by giving the punch lines of the presentation in a simplified and immediately understandable form. The first four sections of this chapter are intended to provide a summary of this type for confabulation theory. Section 1.1 provides background perspective and a nutshell description of confabulation theory. The following three sections then provide a progressively more detailed overview of the human case (with deliberate repetition to aid learning these new concepts). Section 1.5 discusses some of confabulation theory's implications. Finally, Sect. 1.6 provides a brief overview of the book's content.

1.1 In the Beginning

There is strong neuroscience evidence of many kinds suggesting that the initial phase of the story of life on Earth ended about 580 million years ago with a large, rapid, and sustained (to the present) increase in atmospheric oxygen concentration (Canfield et al. 2007, Fike et al. 2006, Kerr 2006). Immediately thereafter, a profusion of macroscopic moving animals emerged (the "Cambrian explosion" of species). The fitness advantages of complex, purposeful movement rapidly drove the evolutionary development of articulated bodies, muscle complements, and the brains and sensory systems needed to purposefully run them.

Movement involves smooth, coordinated control of ensembles of discrete muscles by the animal's brain. Each muscle is supplied with a single neuronal input signal controlling its "analog," continuously variable, level of contraction. Shortly after the emergence of animals capable of sophisticated movement, a new design possibility arose: The extensive neuronal machinery developed to control animal movement could easily be expanded and these additions could be used to control brain *modules*: discrete bodies of neuronal tissue specifically evolved to exploit the pre-existing neuronal muscle-control mechanisms. Instead of conferring motility, these new brain module "movement" processes would carry out a type of information processing called *cognition* or *thinking*. The enormous success of this evolutionary adaptational "redeployment" of movement control led to today's ubiquity of cognition in macroscopic animals (trout, bees, ravens, humans, octopi, et al.). Further, the neuronal mechanisms of cognition were subsequently further adopted as the starting basis for additional brain functions that subsequently evolved, such as the cognitive learning

control system (entorhinal cortex, hippocampus, amygdala, etc.) of mammals. This book concerns itself with explaining the mechanism of thought in detail – with primary focus on the human example.

The purpose of each cognitive module (of which humans have about 4,000 – in contrast with our 700 individual muscles) is to describe one *attribute* that an *object* of the animal's mental universe may possess. This description usually takes the form of *activating* one of a large number of *symbols* (each represented by a small collection of specialized neurons) that are contained within the module. The vast majority of symbols within each module develop during childhood and then remain stable throughout life. Symbols are the fundamental, fixed *terms of reference* that must exist if knowledge is to be accumulated and used over long periods of time.

An individual axonal *knowledge link* (of which the average human adult possesses billions) unidirectionally connects one *source symbol* in one module with one *target symbol* in a second module. These links arise as a result of meaningful causal co-occurrence of the involved pair of symbols (*a la* Donald Hebb). [NOTE: Besides symbol co-occurrence, most animals also impose (e.g., via a centralized cognitive learning control system; as in mammals) the requirement that a new knowledge link also be associated with a reduction in a drive or goal state. Imposition of this requirement has many important advantages – not least of which is the avoidance of a vast buildup of low-value knowledge. Because it is tangential to understanding the mechanism of thought, this "knowledge relevance" requirement and its formidable implementation machinery (it needs to be formidable; because hours often elapse between the temporary establishment of a knowledge link – which the neurons directly involved in implementing the link carry out via instantaneous temporary synapse strengthening – and the realization that this candidate link was involved in a drive or goal reduction) will be ignored in this book. When we need to actually construct knowledge links (e.g., for conducting computer experiments with confabulation), we simply require that all of the knowledge links that are allowed are "of significant value" using some simple criterion. This approach works well for a number of applications – further reinforcing the decision to skip detailed discussion of animal cognitive learning control systems.]

The set of all knowledge links connecting the symbols of one module with the symbols of a second module are collectively termed a *knowledge base* or cortical knowledge *fascicle*. In humans, the set of all cortical knowledge fascicles is, by far, the most massive single brain structure. The capacity for accumulating a vast number of knowledge links is the single most important attribute of the human brain (at an average rate, for most people, exceeding one new knowledge link per second of life); followed by the large symbol capacities of human modules.

Besides implementing symbols, modules also carry out one, and only one, cognitive information processing operation: *confabulation*. Confabulation is the analog of contraction in a skeletal muscle. It occurs only upon receipt of a deliberate *thought command* input to the module. Thought command signals originate

in subcortical structures. Both because not much is known, and to keep the story of confabulation theory focused on cognition, the exact origin of thought commands, and the details of the neuronal processes involved (which involve many subcortical brain nuclei – mostly exactly the same ones as in movement) will be ignored in this book. The origin of the *action commands* that ultimately launch all movement and thought processes (i.e., *behaviors*) **will** be briefly discussed, because they arise as a direct product of cognition (see below).

Strangely, as with a motorneuron signal to a muscle, a thought command is a graded, analog, signal. This is one of several aspects of cognitive information processing that make it starkly alien in comparison with existing concepts such as algorithmic and rule-based computing.

In the milliseconds leading up to a particular target module being commanded to begin (or intensify) a confabulation "contraction," axonal knowledge links from source symbols which are currently *excited* on other selected source modules deliver *input excitation* to neurons representing each knowledge link's target symbol. [The ensemble of modules transmitting excitation are deliberately selected by the overall thought process being executed (thought processes are learned, stored, and recalled in the same manner as movement processes).]

Confabulation is the process of selecting that one symbol (termed the *conclusion* of the confabulation) whose representing neurons happen to be receiving the highest level of excitation. In the case of a single target module undergoing confabulation, this is a simple "winner takes all" competition among the symbols of the target module. At the end of a confabulation all of the neurons which represent the winning symbol are transmitting at high efficacy through any knowledge links that have the conclusion symbol as their source. Through the use of a *neuronal attractor network* circuit contained within the module, a simple confabulation can often be completed in under 100 ms, even if the module implements hundreds of thousands of symbols. Conclusions reached by confabulations in the recent past can be used as the sources of knowledge link input to subsequent confabulations. Conclusion symbols subsequently selected to supply such input are often referred to as *assumed facts* of those subsequent confabulations.

In cognition, single confabulations are rare (much as movements involving contraction of only a single muscle are rare). Usually, thought processes involve an ensemble of tens to hundreds of modules being confabulated contemporaneously during overlapping time intervals – with intercommunication between the symbols of the modules at various points during the gradual, expertly controlled, "contraction" to a single "winning" symbol on each module. This is *multiconfabulation*. A multiconfabulation is typically much more powerful than a single confabulation because it facilitates a process of gradual convergence to a set of "mutually consistent" conclusions; reached by means of mutual communication between the ever-shrinking intermediate sets of candidate conclusions. Multiconfabulation facilitates the application of massive numbers of *relevant* knowledge links (each emanating from a symbol which, at least at that stage of the contraction process, is a viable candidate to be the final conclusion of that

module). Properly executed, a multiconfabulation allows multiple opportunities to "cross-check" the lists of not-yet-eliminated candidate symbol conclusions to ensure that the final conclusions reached (collectively termed the *confabulation consensus*) are mutually consistent with respect to the available knowledge. Thus, the slowed convergence process of multiconfabulation (with the rising "contraction" thought command signal corresponding to the gradual shrinking of the list of remaining candidate conclusions from which the final single winning symbol will be selected) is an essential aspect of cognition. An information processing system employing carefully and skillfully coordinated smooth information processing (thought) commands to the involved processors (modules) is starkly alien in comparison with all existing concepts of information processing.

How can confabulation – a simple competition process between the symbols of a module on the basis of which symbols are receiving the most axonal excitation – be the complete and final explanation for all aspects of cognition? This would seem to imply that, in some sense, confabulation is a powerful, general purpose, universal decision-making procedure. Surely there must be some new and powerful mathematics underlying it. And there is. Describing and characterizing this surprising and strange cognitive mathematics is a main focus of this book.

Finally, a key unanswered neuroscience question is the origin of *behavior* (thought processes and movement processes). Obviously, animals launch many behaviors every minute – often many per second. There must be a unified source of these actions. The shockingly simple answer is that every time a confabulation is completed, action commands, uniquely associated with the winning conclusion, are instantly launched and sent to subcortical structures (e.g., the basal ganglia) for evaluation and, perhaps, execution. All non-reflexive and non-autonomic behavior originates in this manner.

The axonal associations between each symbol in a module and its fixed set of action commands are termed *skill knowledge*. While skill knowledge is stored in cerebral cortex, it is established and modified by subcortical brain nuclei. Skill knowledge is very different from cognitive knowledge – e.g., far from being very long lasting like cognitive knowledge, skill knowledge, if unused, fades rapidly – often within a few weeks. Skill knowledge is "use it or lose it." Also, skill knowledge is inherently "overwritable," allowing more recent skill practice session performances to "overwrite" older, presumably less competent, skill knowledge. In order to remain focused on cognition, very little is said in this book about skill knowledge.

A major advantage of cognition is that all cognitive knowledge is *interoperable*. The knowledge links delivering excitation to a particular thought process might emanate from symbols representing auditory, visual, linguistic, or even movement process attributes of mental world objects. The type of attribute that their source symbols encode makes no difference: the knowledge link excitation input to the symbols of the involved target modules are simply approximately summed up. To appreciate the power of this capability, consider the difficult challenge faced by an algorithmic information processing researcher who is

attempting to combine image and sound data from a theatrical motion picture to accurately recognize specific movie actors.

Cognition is a "core competence" of macroscopic multicellular Earth animals. In each taxonomic category, species with particularly high cognitive skill stand out: bees among insects, humans among primates, jays and ravens among birds, cetaceans among aquatic mammals, etc. Anthropocentrism puts humans on the highest pedestal; but all cognitive "champions" have their distinctive relative superiorities. I leave it to philosophers, SETI researchers, future interstellar explorers and theologians to incorporate the insights of confabulation theory into larger points of view and to address sweeping universal questions (such as: Is confabulation the unique extant approach to natural intelligence in our universe, or are there others?). This book concentrates on the confabulation theory explanation for human cognitive function and on the use of confabulation theory as the basis for building intelligent machines.

1.2 Cerebral Cortex and Thalamus: The Seat of Cognition

There is strong neuroscience evidence of many kinds suggesting that the "information-processing" involved in all aspects of cognition (seeing, hearing, planning, language, reasoning, control of movement and thought, etc.) is carried out by the cerebral cortex and thalamus. There is also strong evidence that the "cognitive knowledge" used in this processing is stored in the cerebral cortex. Beyond vague statements of this sort, at present essentially nothing is known about how cognition (which will also be referred to in this book as *thinking*) works, or about what cognitive *knowledge* is.

This book presents the first concrete and detailed (and thus falsifiable) scientific theory of how thinking works. This *confabulation theory* proposes the specific neuroanatomical structures, and their functions, that are involved in human cognition.

The two main human neuroanatomical structures postulated by confabulation theory to be involved in the implementation of thought are *thalamocortical modules* (Fig. 1.1) and *knowledge bases* (Fig. 1.2). These structures, which constitute the "information-processing hardware" used to carry out thought, exist within the cerebral cortex and thalamus. The human brain possesses roughly 4,000 thalamocortical modules and roughly 40,000 knowledge bases[2]. All vertebrates (and even invertebrates such as bees and octopi) are postulated to possess functionally analogous structures, albeit in smaller quantities.

[2] For concreteness, confabulation theory specifies many numeric values quantifying aspects of the theory's postulated human neuroanatomical structures. These can be thought of as crude, rough order of magnitude, estimates of means; with most quantities also having significant variance. For simplicity, value accuracy and variability are not discussed.

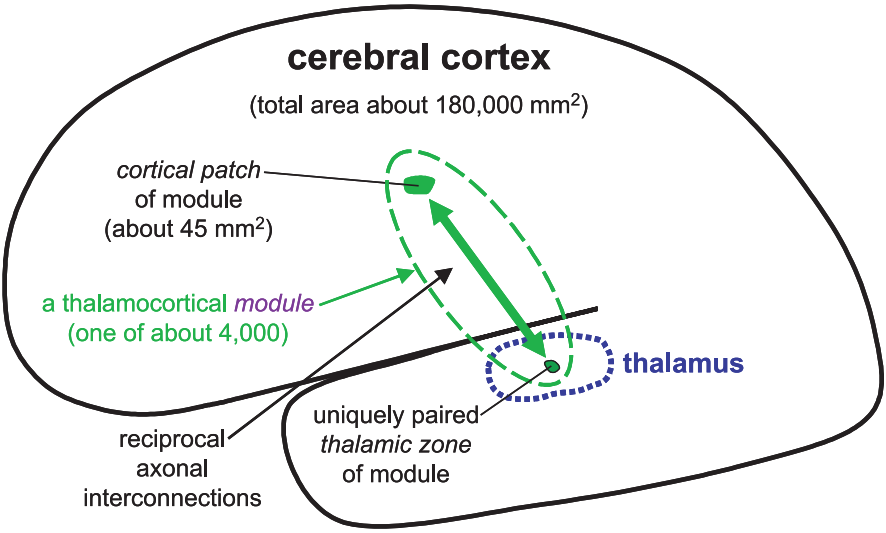

Fig. 1.1. A *thalamocortical module* (one of roughly 4,000 in the human brain). Each thalamocortical module is comprised of a small *patch* of cerebral cortex and a uniquely paired small *zone* of thalamus. The cortical patch of each module is reciprocally axonally connected with the thalamic zone of the module. The cortical patches of different modules are largely disjoint (partial overlaps do likely occur). Similarly for their thalamic zones. The union of the cortical patches of all thalamocortical modules comprise the entire area of cerebral cortex. However, the union of the thalamic zones of all modules do not comprise all of the thalamus

The cortical neural tissue encompassed by each thalamocortical module bears resemblance to that of the "cortical columns" proposed decades ago (Mountcastle 1988; Paxinos and Mai 2004), except that the cortical component of a module is roughly 200 times larger in volume than a cortical column. The postulated functions of thalamocortical modules are also completely different from those envisioned for columns.

Knowledge bases are related to the axonal links between pairs of cortical "neuron populations," as postulated vaguely by Hebb 57 years ago (Hebb 1949) and more concretely and recently by Abeles (Abeles 1991).

The level of description of function offered by confabulation theory is one level up from that of the individual neurons. The study of how these functions are implemented at the neuron and molecular levels is termed *confabulation neuroscience*. Since very little is known, the discussion of confabulation neuroscience in this book (principally Chaps. 2, 3, 5, and 8) is mostly speculation, and will likely require significant revision as more is learned.

As noted in bibliographic citations throughout the book, and discussed explicitly in Chaps. 3, 5, and 8, confabulation theory is strongly related to many bodies of past research.

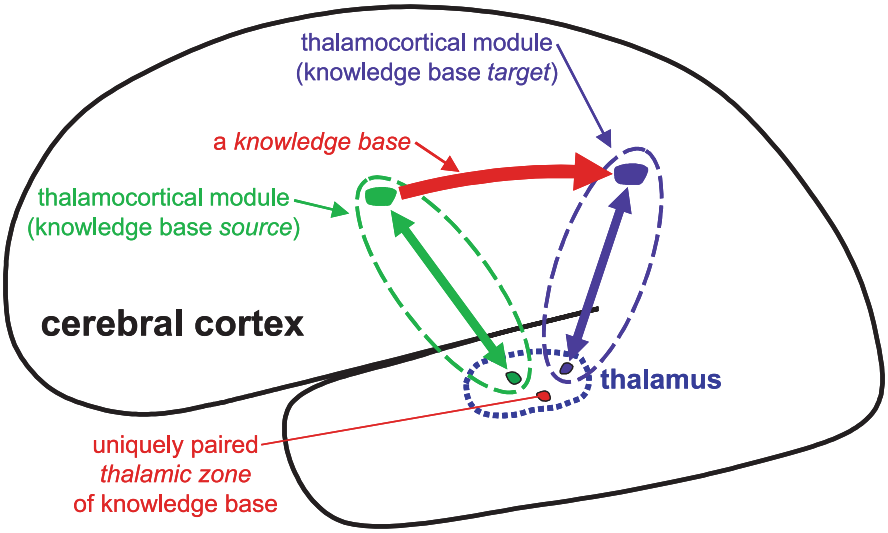

Fig. 1.2. A cognitive *knowledge base* (one of roughly 40,000 in the human brain). Roughly 40,000 ordered pairs of thalamocortical modules (*source* and *target* modules) are selected (by genetically specified developmental processes carried out in childhood) to each have their cortical patches unidirectionally linked by a knowledge base. Each knowledge base is comprised of a large number (often millions) of individual *knowledge links*. Much like a thalamocortical module, each knowledge base is postulated to be paired with a unique, dedicated zone of thalamus which is postulated to be involved in that knowledge base's functional *enablement*. The combination of the thalamic zones of the modules and knowledge bases make up the vast majority of the thalamus

1.3 The Four Key Elements of Confabulation Theory

Today, the cognitive information-processing and cognitive knowledge acquisition, storage, and use functions of cerebral cortex and thalamus are completely unknown. Confabulation theory specifies them completely. In particular, confabulation theory postulates four key functional elements (#s 1, 3, and 4 implemented by thalamocortical modules and #2 implemented by knowledge bases) which together comprise *the neuronal information-processing "hardware" of thought*. These four key elements, and the manner in which thalamocortical modules and knowledge bases implement them, are each individually sketched in the four sub-sections of this section. The manner in which these functional hardware elements are used to implement thought is explored in detail in the book's video presentation (and the associated presentation notes) and in Chaps. 3, 4, 6, and 7.

1.3.1 Confabulation Theory Key Element #1: Each Thalamocortical Module Describes One Mental Object Attribute

Each thalamocortical module (Fig. 1.3) is used for describing one *attribute* which an *object* (sensory, language, abstract, movement process, thought process, plan, etc.) of the mental universe may possess. To describe its attribute, the module is equipped with a large collection of *symbols*. When utilized for describing an object, a module typically *expresses* one symbol chosen from its collection. The

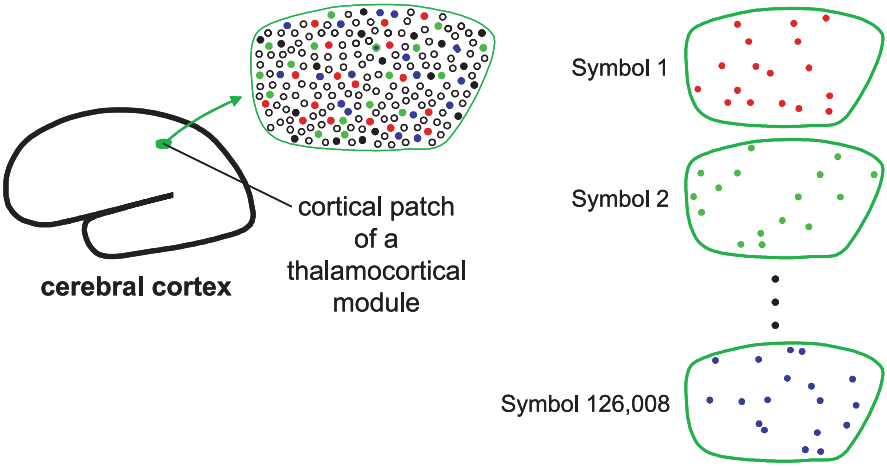

Fig. 1.3. A primary function of each thalamocortical module is to describe exactly one *attribute* that an *object* of the mental universe (a sensory object, a motor process object, a thought process object, a plan object, a language object, etc.) may possess. To carry out this object – attribute – description function, each module implements a large collection of *symbols*. When utilized for describing an object, a module typically *expresses* one symbol chosen from its collection (primary sensory and motor modules usually express multiple symbols). Each symbol is represented by roughly 60 neurons selected (approximately uniformly at random) from a special population of *symbol-representing neurons* (shown as colored dots within the enlarged depiction of the module's cortical patch) that reside within the cortical patch of the module. Here, a module with 126,008 symbols is depicted. Each symbol's subset of 60 neurons is shown schematically. Symbols are mostly formed in childhood and then remain stable throughout life – they are the *stable terms of reference* that must exist if knowledge is to be accumulated across decades. The famous *binding problem* (von der Malsburg 1981) does not apply to confabulation theory because each of the attribute description symbols of an object is typically linked to many of the others pairwise by *knowledge links* (see Sect. 1.2.2). In effect, a mental world object *is* any reasonably large subset of its pairwise-linked attribute description symbols. Thalamocortical module symbol sets (the collection of different descriptive terms for representing the object attribute that the module is responsible for encoding) are the first of the four key functional elements of confabulation theory

symbols of a module are mostly created in childhood and are stable over decades. Symbols are the *stable terms of reference* which must exist if knowledge is to be accumulated over long periods of time. For example, in a human a particular thalamocortical module might be responsible for representing the *name* of an object. This module might possess 128,008 symbols, representing words, phrases, and punctuations such as: `mother`, `father`, `President Kennedy`, `Bunsen burner`, `lunar regolith`, `candy`, and `Candy`.

1.3.2 Confabulation Theory Key Element #2: Knowledge Links Connect Pairs of Co-occurring Symbols

Although the concept of cognitive human knowledge – something which is acquired, stored, and then used – has been in widespread use for millennia, even today there is no understanding of the mechanisms involved (other than the persistent suspicion that Hebbian synaptic modification might somehow be involved) or of the nature of knowledge. Confabulation theory (see Figs. 1.4 and 1.5) specifies precisely what cognitive knowledge is, how it is acquired, how it is stored, and how it is used in thinking (Sect. 1.3.3).

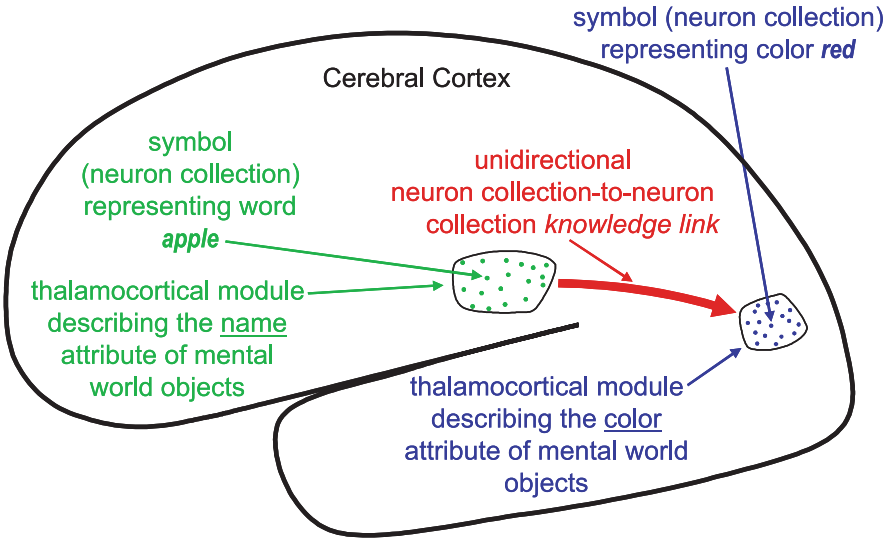

Fig. 1.4. A cognitive *knowledge link*. Here, a human subject is viewing and considering a red apple. A visual thalamocortical module is expressing a symbol for the color of the apple. At the same time, a language thalamocortical module is expressing a symbol for the name of the apple. Pairs of symbols which *meaningfully co-occur* in this manner have unidirectional axonal links, termed *knowledge links* (each considered a single *item of knowledge*), established between them via synaptic strengthening (assuming that the required axons are actually present – this is determined by genetics). The average adult human has billions of knowledge links, most of which are established in childhood. The rate of human knowledge acquisition often exceeds one link per second of life

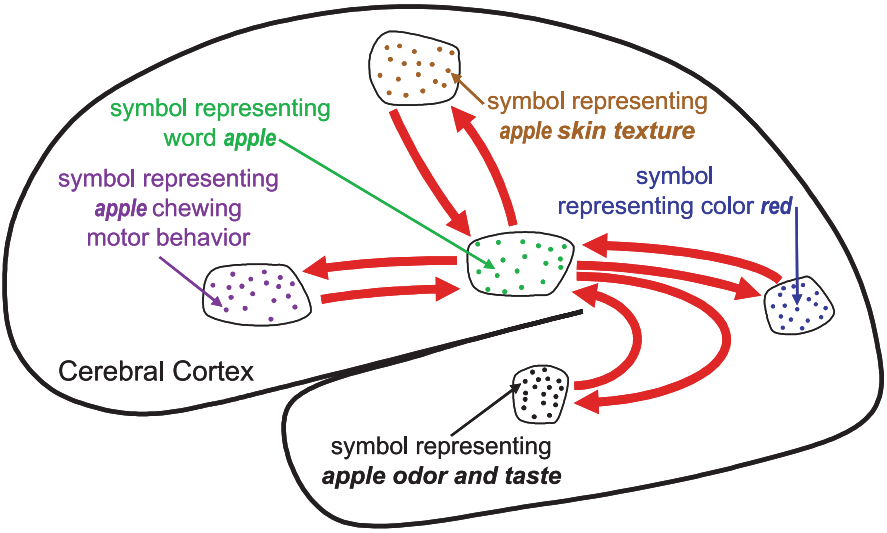

Fig. 1.5. Billions of pairs of symbols are connected via knowledge links. The set of all knowledge links joining symbols belonging to one specific *source* module to symbols belonging to one specific *target* module is termed a *knowledge base*. In the human brain, knowledge bases take the form of huge bundles of axons termed *fascicles*, which together make up a large portion of each cerebral hemisphere's ipsilateral white matter. Each module also typically has a knowledge base to its contralateral "twin" module (and perhaps to a few others near its twin) – which together constitute the *corpus callosum* fascicle linking the two cerebral hemispheres. Here, reciprocal knowledge links (red arrows), only some of which are shown, connect each expressed symbol representing an attribute of an apple pairwise with other such symbols. When an apple is currently present in the mental world, it *is* its collection of knowledge-link-connected symbols which are currently being expressed. There is no binding problem because all of these symbols are mutually "bound" by their previously established pairwise knowledge links. Shockingly, confabulation theory contends that such knowledge links – formed exclusively on the basis of meaningful symbol pair co-occurrence – are the only type of knowledge used (or needed) in cognition! Knowledge links are the second of the four key elements of confabulation theory

1.3.3 Confabulation Theory Key Element #3: Confabulation – The Information-Processing Operation of Thought

The vague notion that cognition employs some sort of "information-processing" has been around for millennia. Today, the understanding of the exact nature of this "cognitive information-processing" is roughly the same as it was in 350 B.C. – the time of Aristotle (arguably the first neuroscientist). Confabulation theory states explicitly and exactly that cognition involves only one information-processing operation – *confabulation* (see Fig. 1.6): a simple winners-take-all

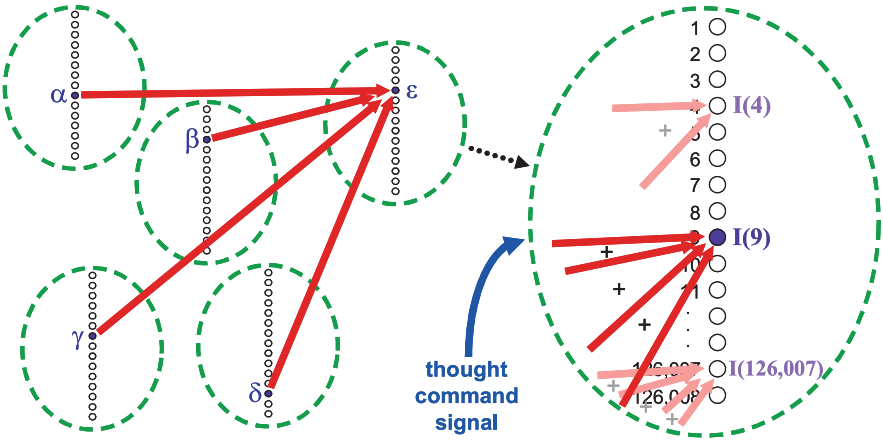

Fig. 1.6. *Confabulation* – the only information-processing operation used in cognition. Here, a concrete example involving five thalamocortical modules is shown (for simplicity, each module is illustrated as a dashed green oval with a list of that module's symbols inside it). See text for details. Confabulation is the third of the four key elements of confabulation theory

competition between symbols on the basis of the total input excitation they are receiving from knowledge links.

As seen in Fig. 1.6, the four modules on the left are each describing the attributes of one or more mental world objects by each expressing a single symbol: α, β, γ, and δ. Each of these four expressed symbols has a large number of knowledge links connecting it with symbols of the fifth module (of which four knowledge links, linking each expressed symbol to symbol ε of the fifth module, are shown). The situation within this fifth module, which is about to undergo *confabulation*, is shown enlarged on the right. For illustration, symbol 4 of this module is receiving two knowledge links (one from symbol α, and one from symbol γ), whereas symbols 9 and 126,007 are receiving knowledge links from all of α, β, γ, and δ. Each knowledge link is delivering a certain quantity of *input excitation* to the neurons of its target symbol.

The input excitations arriving at symbol k from different knowledge links are summed to yield the *total input excitation for symbol k*: I(k) (this summation is noted by the plus signs between the knowledge links in the enlarged illustration of module five). [As discussed extensively in this book, this *additive knowledge combination* property is one of the paramount reasons for the enormous information-processing power and flexibility of thought.].

Upon being commanded to do so (by a deliberate externally supplied *thought-command signal* – analogous to the motorneuron input to a muscle – illustrated by a blue arrow in Fig. 1.6), the symbols of the fifth module compete with one another (via a highly parallel, fast, *neuronal attractor network* function), yielding a final state in which all of the neurons representing the symbol with the largest

input intensity I (in this example, symbol 9) are highly activated and all other symbol-representing neurons are not. This "winners-take-all" information-processing operation is called *confabulation*, and the winning symbol is termed the *conclusion*.

Confabulation is hypothesized to be the only information-processing operation involved in thought. In the Fig. 1.6 example, there is only one confabulation taking place. Ordinarily, confabulations on multiple modules take place together, with convergence to the winning symbol slowed somewhat to allow mutual interaction during convergence ("comparing notes" in order to arrive at a *confabulation consensus* of final conclusions). In such a *multiconfabulation*, often millions of items of knowledge, each emanating from a viable candidate conclusion, are employed in parallel in a "swirling" convergence process. (As discussed extensively in this book, this is another paramount reason for the enormous information-processing power and flexibility of thought.) Confabulation is the third of the four key elements of confabulation theory.

Confabulation is starkly alien in comparison with existing concepts in neuroscience, computational intelligence, neural networks, computer science, AI, and philosophy in general. For example, computer CPUs all follow the Turing paradigm: when commanded via a specific, digital, instruction code they execute a pre-defined logical or arithmetic instruction on specified variables. Thalamocortical modules, on the other hand, have only one information-processing "instruction" – confabulation. Further, the command to confabulate (termed the *thought-control command* – which is delivered to the confabulating module from outside cerebral cortex and thalamus) is not digital; rather, it is *analog*. Yet the result of a completed confabulation is digital: a single symbol. Very weird.

The ultimate challenge is to show that it is possible to explain Newton, Mozart, Einstein, and Crick using confabulation. That will probably take a while. Yet, the evidence presented in this book is intended to build confidence that this challenge will someday be met.

1.3.4 Confabulation Theory Key Element #4: The Conclusion → Action Principle – The Origin of Behavior

One of the most obvious aspects of brain function (and therefore one of the most consistently ignored) is that animals typically launch many behaviors every second they are awake. Most of these are *microbehaviors* (small corrective modifications to ongoing behaviors), but, typically, many times per hour major new behaviors are launched, predicated on newly emerged events. Beyond simple reflexes (e.g., knee jerk) and autonomic reactions (e.g., digestion), no understanding of how and why behaviors originate currently exists.

Confabulation theory proposes the *conclusion → action principle* (Fig. 1.7), which states that every time a confabulation operation on a thalamocortical module reaches a conclusion, an associated set of *action commands* are launched from the cortical patch of the module via axons which proceed towards sub-cortical

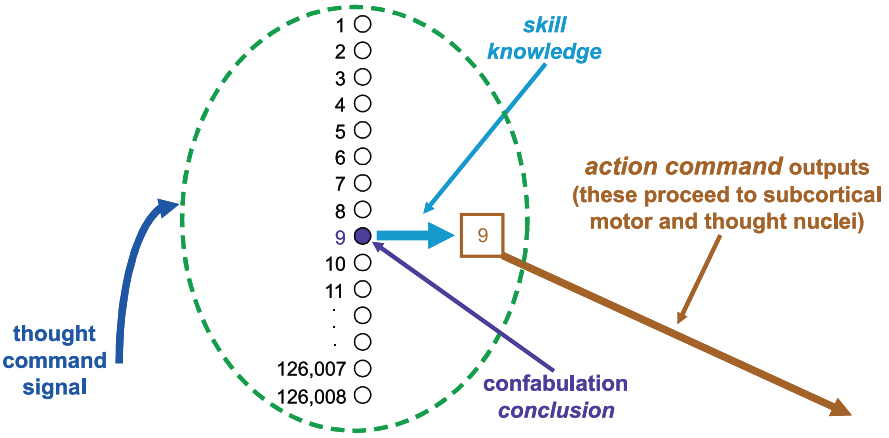

Fig. 1.7. The *conclusion → action principle*: hypothesized to be the origin of all non-reflexive and non-autonomic behavior. Here, a thalamocortical module (illustrated, in consonance with Fig. 1.6, as an abstract "oval" structure containing a list of the module's symbols) has successfully completed a confabulation operation (under control of its externally supplied thought-command signal) and reached a conclusion (symbol number 9 as in Fig. 1.6). Whenever a module completes a confabulation and reaches a conclusion it immediately causes a set of *action command* outputs to be launched (these outputs proceeding to sub-cortical nuclei). The specific action command outputs that are launched are those which have been previously *associated from* this specific conclusion symbol via a completely separate, sub-cortically managed, *skill-learning* process. These action command outputs can cause behaviors to occur. The conclusion → action principle is the fourth and last of the key elements of confabulation theory

structures. Often, these action commands lead to the initiation of *behaviors* (either immediately or after further evaluation). All non-reflexive and non-autonomic behavior arises in this manner.

Action commands can be regarded as *suggested behaviors* – which sub-cortical structures either immediately execute, consider further for future execution, or (e.g., if the suggested behavior is not consistent with past successful reductions in currently elevated goal or drive states) discard.

The *associations* between each symbol of a module and the specific action commands which are to be issued when that symbol wins a confabulation competition are termed *skill knowledge*. Skill knowledge is formed via selective strengthening of special synapses within cerebral cortex; but the involved *skill-learning* process is controlled by sub-cortical structures.

Skill knowledge, although implemented by synapses in cortex, differs greatly in neuroanatomical location and physiological properties from cognitive knowledge links. For example, unlike a cognitive knowledge link (which, if solidified over the 100 hours following the initial symbol pair co-occurrence, is extremely durable), skill knowledge is often fragile and short-lived (this is important for *rehearsal*

learning of skills, where later, more competent skill knowledge needs to "supersede" and supplant earlier, less perfected skill knowledge).

Behavioral triggering, skill knowledge, and skill learning are not parts of thinking (they come into play only after thinking has completed its job of reaching conclusions) and so they are not discussed much in this book. Of course, this decision is subject to the criticism that thinking itself is utterly dependent upon the *thought-command sequences* which control the operation of the thalamocortical modules involved in a particular thought process. These thought-command sequences are learned, stored, and recalled in exactly the same manner (using knowledge links) as the movement command sequences (actually, *postural goal* sequences) employed in movement. So, thought begets movement and thought (both termed *actions*) in an endless action – thought – action – thought – action – thought – … sequence during wakefulness (thereby exorcising the homunculus hiding behind a curtain pulling the control levers of the brain and body). Actually there is quite a bit that could be said about all this; but this topic is deferred to a future edition. In this book, the focus is on the basic mechanism of thought.

1.4 Cognitive Brain "Hardware" and "Software"

The four key functional elements of confabulation theory described in Sect. 1.3 constitute the "information-processing hardware" upon which confabulation theory contends thinking is implemented. But what about the "software" of thought (the procedures, called *thought processes*, for using the hardware)?

A central hypothesis of confabulation theory is that thinking is a phylogenetic outgrowth of movement. Animals began moving over a billion years ago. The mechanisms for flexible, adaptive control of movement emerged early and expanded rapidly. As animal movement complexity and capability grew, a new design possibility emerged: the elaborate machinery already developed for controlling movement could be applied to brain tissue. In particular, discrete brain structures, *modules*, emerged that could be controlled exactly like individual muscles. By manipulating these modules in properly coordinated "movements" (thought processes), information-processing could be carried out – thereby further enhancing competitive success.

As discussed in Sect. 1.3.3, each human thalamocortical module has a single thought-command input signal that tells it when to "contract." This is analogous to the roughly 700 muscles of the human body, each of which has a single input signal (motorneuron input) that commands it to contract. Just as with a muscle, the thought-command input to a module is an *analog* signal: it can range from a low level ("contract a little") to a highest level ("contract with maximum force"); where "contraction" corresponds roughly to the rate of convergence, from multiple candidate conclusions to a single conclusion, of a module's confabulation competition.

In effect, the human brain thinks by maneuvering subsets of 4,000 digital processors (the thalamocortical modules) through smooth, graceful, thought maneuvers. These thought processes are learned, stored, and recalled just as movement processes are learned, stored, and recalled. At higher hierarchical levels, closely related movement processes and thought processes are often stored mixed together in the same knowledge links.

Just as the repertoire of human movement can be vast (walking, writing, running, cartwheels, uneven parallel bar routines, pole vaulting, etc.), so the repertoire of thought can contain a vast variety of different ways of using thalamocortical modules. However, at the present time, confabulation theory has only identified a few of these ways. And only two of these, a single isolated confabulation (crudely analogous to flexing of a single muscle) and *multiconfabulation swirling* (crudely analogous to walking – the most basic and useful of human movements), have received significant study. All of the remaining chapters of this book discuss these two types of basic thought process.

As is discussed in detail in the video presentation, brains carry out a multitude of functions in addition to cognition. Quite a few of these interact intimately with, and are required to implement, thought processes. However, these other brain functions are poorly understood and are only briefly mentioned in this book. The thought processes considered here (single confabulations and multiconfabulations) are implemented using an *external thought controller* executing a crude, contrived thought process. The only feedback that a *thought controller* gets from the thought process being executed on the involved collection of modules is knowledge of when a module has reached a conclusion (in effect, an action command output, as in Sect. 1.3.4). This feedback can be used to trigger recall and playback of a different "canned" thought process. While this approach only implements a tiny subset of the capabilities of real brain thought and movement control, as the reader will see, it is still possible to achieve interesting results.

1.5 Implications of Confabulation Theory

Confabulation theory has a variety of implications. A few examples are discussed here.

Since all of cognition is "categorical" (i.e., based upon the symbol sets of the thalamocortical modules), the total number of modules, and the number of symbols in each of those modules, provides a reasonable estimate for the "descriptive power" of a brain. A trout may have only a few tens of modules, each with a few hundred symbols. A raven might have hundreds of modules, each with many hundreds of symbols. A human probably has thousands of modules, each with thousands to hundreds of thousands of symbols. Similarly, the total number of knowledge links that an animal possesses gives a crude quantification of how "smart" that animal is (although, clearly, the distribution of those knowledge links also matters: idiot savants may have huge numbers of knowledge links).

The experiments of Chap. 6 imply that the average human possesses billions of items of knowledge, of which the majority are often obtained in childhood. Some humans may possess tens, or perhaps even hundreds, of billions of items of knowledge. Clearly, since there are only about 32 million seconds in a year, the average rate of knowledge acquisition often exceeds one item per second and might sometimes exceed 100 items per second. It is thus not surprising that we need to sleep a third of the time in order to catch up with evaluating and selectively solidifying each day's new cognitive knowledge links (i.e., implement cognitive learning control decision-making for recently established, and intrinsically rapidly fading, temporary knowledge links – which is probably the main activity of sleep).

Humans (and animals in general) are almost certainly much "smarter" than has been generally appreciated. Assuming such findings are confirmed, fields as diverse as psychology, education, philosophy, psychiatry, medicine (both human and veterinary), law, and theology will need to be extensively overhauled.

With one relatively small exception, the axonal connectivity between the thalamocortical modules in the human brain seems to roughly resemble that of other great apes. That one exception is the modules of the human language faculty – which seem to connect widely to modules in many other faculties. In this sense, language is the *hub* of human cognition. It seems likely that this (along with having a brain which is, overall, over three times larger) can explain some of the commanding power of human thought in comparison with that of other apes. As we learn more about cetaceans, it may well be that some of them (and perhaps other species as well, such as jays, ravens, and parrots) also have this *language hub cognitive architecture* characteristic to some degree.

The near-term implications of confabulation theory for neuroscience are uncertain. Neuroscience is dominated by bottom-up thinking and by "methods." To succeed, neuroscientists must often spend the decade after completing their Ph.D. developing their own effective experimental methods. The subset of aspirants who successfully complete this process must then, in general, inaugurate and manage a large lab that quickly acquires enormous built-in inertia. After completing this arduous initiation at about age 40, few of these newly established neuroscientists are going to be interested in abandoning, or significantly altering, their research direction in order to begin to follow up on the hypotheses of confabulation theory. Thus, integration of confabulation theory into neuroscience is likely to be largely confined to new investigators who decide to pursue experimental exploration of confabulation theory's neuroscience implications (probably mainly using human subjects carrying out controlled thought processes while being monitored by brain activity scanners with greatly improved spatial and temporal resolution). Assuming this established social pattern continues to hold, it seems unlikely that confabulation neuroscience can join the mainstream of the subject until the next decade.

Notwithstanding the above, members of the small community of mathematical neuroscientists may soon realize that, given the hard constraints provided by confabulation theory, it may be possible to tackle large-scale understanding of

brain function. For example, it may be possible within a few years to build an integrated functional mathematical model of cerebral cortex, thalamus, basal ganglia, subthalamus, red nucleus, substantia nigra, hippocampus, amygdala, hypothalamus, spinal cord, locus coeruleus, pons, and cerebellum. This model may well answer most of the large questions of neuroscience that remain after confabulation theory.

A large-scale human brain modeling project of this sort will surely require a widely knowledgeable and exceptionally well educated team of hundreds of mathematical neurobiologists and computer scientists operating as willing and compliant subordinates under the hierarchical command of a master genius. The usual "herd of cats" sort of scientific research program would probably not work effectively in this instance. I personally know at least five people who could each probably successfully lead such an effort. Such an integrated brain modeling project is, in my opinion, one of the most important tasks that the human species should now carry out. It will be expensive (probably exceeding $200,000,000 per year for a decade; along with another $400,000,000 for a proper building to house the project and the budget for the required equipment). A single, open, international project of this type would seem ideal. However, given the potential economic and national security implications, multiple projects of this type seem more likely. With respect to these practical implications of confabulation theory, I leave it to you, the reader, to form your own opinion as you absorb the book's content.

1.6 Content of the Book

The content of the eight chapters and two DVDs of this book is briefly surveyed below:

- **Chapter 1: Introduction**
  An introductory overview of confabulation theory: comments on some of the theory's possible implications and presentation of this overview of the book's contents.

- **Chapter 2: Video Presentation Viewcells**
  The viewcells used in the book's DVD video presentation are presented. To help with understanding and retention of the material, each of these should be referred to while it is being presented during the video.

- **Chapter 3: The Mathematics of Cognition**
  An introduction to the mathematics of confabulation theory. Comments on the relationship between cogency maximization and Bayesian analysis. An extensive discussion of the status of confabulation neuroscience. Comments on the origins of confabulation theory.

This chapter is based on the original publication

> Hecht-Nielsen R (2006) The mathematics of thought. In: Yen GY, Fogel DB (eds) Computational intelligence: Principles and practice. IEEE Computational Intelligence Society, Piscataway, NJ, pp 1–16

and is adapted here in accordance with IEEE copyrights.

- **Chapter 4: Cogent Confabulation**
  Mathematical foundations of confabulation theory are presented, including statement and proof of the Fundamental Theorem of Cognition and the theorem showing that cogent confabulation within a logical information environment yields Aristotelian logic. Computer experiments with a single confabulation are presented, with all details provided. Replication of these single confabulation experiments is the logical starting point for those wanting to gain hands-on experience with confabulation architectures.
  This chapter is a reformatted reprint of the original publication

  > Hecht-Nielsen R (2005) Cogent confabulation. Neural Networks 18:111–115, Copyright (2005)

  used with permission from Elsevier.

- **Chapter 5: Confabulation Neuroscience I**
  A concise overview of confabulation neuroscience. This material is prerequisite for Chap. 6.
  This chapter is based on the original publication

  > Hecht-Nielsen R (2006) The mechanism of thought. In: Proceedings of the World Congress on Computational Intelligence. 16–21 July, Vancouver, BC, Canada. IEEE Press, Piscataway, NJ

  and is adapted here in accordance with IEEE copyrights.

- **Chapter 6: The Mechanism of Thought**
  Computer experiments with multiconfabulation are presented, with all details. These *sentence continuation* experiments illustrate that thinking is exactly like moving. Replication of these multiconfabulation experiments is the second logical step for those wishing to gain hands-on experience with confabulation architectures.

- **Chapter 7: Mechanization of Confabulation**
  Further details of confabulation architecture design and implementation are presented. Approaches for application of confabulation architectures to language, vision, and hearing are discussed in some detail.
  This chapter is based on the original publication

  > Hecht-Nielsen R (2006) The mechanization of cognition. In: Bar-Cohen Y (ed) Biomimetics. CRC Press, Boca Raton, FL, pp 57–128

  and is adapted here from the original with kind permission of the publisher.

- **Chapter 8: Confabulation Neuroscience II**
  An expanded discussion of confabulation neuroscience.
  This chapter is based on the Appendix of the original publication

  Hecht-Nielsen R (2006) The mechanization of cognition. In: Bar-Cohen Y (ed.) Biomimetics. CRC Press, Boca Raton, FL, pp 57–128

  and is adapted here from the original with kind permission of the publisher.

- **DVDs**
  The book's two DVDs (attached to this book) contain the following material:

  1. *The Mechanism of Thought* video presentation (Part I on DVD Disk 1 and Part II on DVD Disk 2).
  2. PDF file of the *Viewcells* used in *The Mechanism of Thought* video presentation. This computer-readable file is included on both Disk 1 and Disk 2.
  3. PDF file of the *Presentation Notes* for *The Mechanism of Thought* video presentation. This computer-readable file is included on both Disk 1 and Disk 2. [Note: These *Presentation Notes*, intended for use as courseware, are probably the most important component of the book.]