Chapter 1. Introduction

The current project is a culmination of past researches by Richard B. Wells on the study of function of mind [Wells, 2011f] (Fig.1.1). The researches come under the umbrella term 'Martian Program" whose goal is to discover functional system architecture minimally required to model the development of sensorimotor intelligence in first months of an infant's (human) life. Thus the Martian is an abstract proxy neural network system.

The first Martian Program spanned from 2006 to 2009. From the functions the Martian could and could not exhibit it was concluded that an anatomical-functional model was inadequate and a "mind-model" was required. Wells concluded that a better functional knowledge of the neural-code was required [Wells, 2011a; Wells, 2011b; Wells, 2011c; Wells, 2011e].

Developmental psychology has shown many psycho-functional aspects in growth of early sensorimotor intelligence [Piaget, 1952b]. The Martian is intended to exhibit these abilities. This came in concert with culmination of Wells' research on the epistemological foundations of phenomenon of mind which began from mid-1990 [Wells, 2006]. Based on these foundations a mathematical theory of mind function was developed called "mental physics" [Wells, 2009].

Based on mental physics, the psycho-functional abilities at sensorimotor stage begins with the ability to represent appearances as intuitions. Since Martian-I lacked this ability, Martian-II was introduced in 2011 [Wells, 2011f]. The second Martian program incorporates mental physics and hence a mathematical ability to represent intuitions called mathematical sensibility.

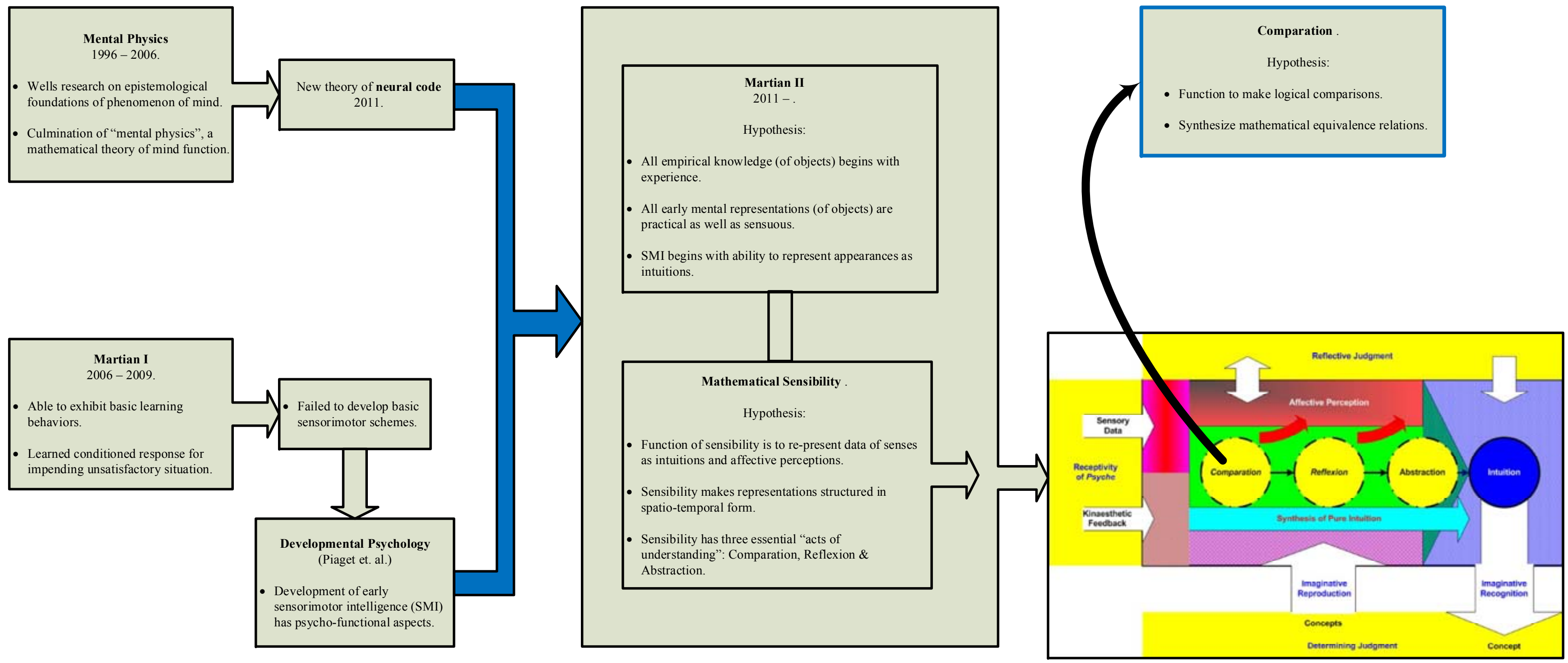


Figure 1.1. Roadmap connecting past research on the Martian and mental physics to current Martian-II program. Comparation as seen in the box illustrating synthesis of sensibility is one of the three "acts of understanding", the Verstandes Actus.

Problem Statement

Mathematical sensibility makes representations (parástases) structured in spatio-temporal form by re-presentations of data of senses to form of intuitions and affective perceptions. This is done from three basic "acts of understanding": **Comparation**, **Reflexion** and **Abstraction** (Fig.1.1). It should be noted that sensibility does not judge. This task is performed by reflective judgment.

The current project is the comparation project. The role of comparation in **sensibility** is to make logical comparisons. Thus its function is to synthesize mathematical equivalence relations. The aims of the project are therefore:

- Discover minimal neural network anatomy to generate equivalence relations.
- Determine proxy functions for interaction between sensibility and reflective judgment.
- Determine proxy functions for generating equivalence relations.
- Demonstrate functional comparation in an embedding field neural network.

Further discussions on the problem of logical comparison and approaches taken for this project is discusses in the succeeding chapters.

A brief overview of general history of our human journey in understanding 'how the brain works?' is given below. Nobody has tackled the problem of logical comparison in the study of mind based on mental physics developed by Wells. However, there have been past and contemporary studies on problems related with discriminating equivalences. These are discussed after the general history overview. Since an exhaustive description is beyond the scope and for the sake of clarity the general discussion is divided in terms of anatomical-physiology and philosophical ideas and arguments. Beginning with the former a brief history is illustrated in figure 1.2.

Brain Function

Alexandrian period saw a major change in idea about brain function from the past concept of brain as repository of soul [Corner, 1919]. The three major figures were Herophilus (~300 B.C), Erasistratus (~260 B.C) and Galen (129 – 199 A.D). Herophilus was first to consider ventricles (cells of the brain) as location for mental processes, favoring 4^{th} ventricle as the precise site [Clarke & O'Malley, pp.713-714, 1968]. This view of ventricular localization shifts from popularity to losing favor back-and-forth until the 19^{th} century. Thus it was probably the longest surviving biological thought [Clarke & Dewhurst, pp.85, 1972]. Erasistratus considered brain convolutions (gyri) being proportional to intelligence and gave the analogy 'coils of small intestine' for their appearance [Clarke & O'Malley, pp.631, 1968].

Galen, the great anatomist of antiquity who gave the detailed description of the ventricles also identified and traced motor nerves to cerebellum and sensory nerves to the cerebrum [Clarke & O'Malley, pp.460-461, 1968]. He considered intellectual functions to have three constituents; imagination, reason and memory. He preferred to locate them to brain substance and not the ventricles but avoided their precise location [Clarke & Dewhurst, pp.10, 1972]. Galen's model is illustrated in figure 1.3.

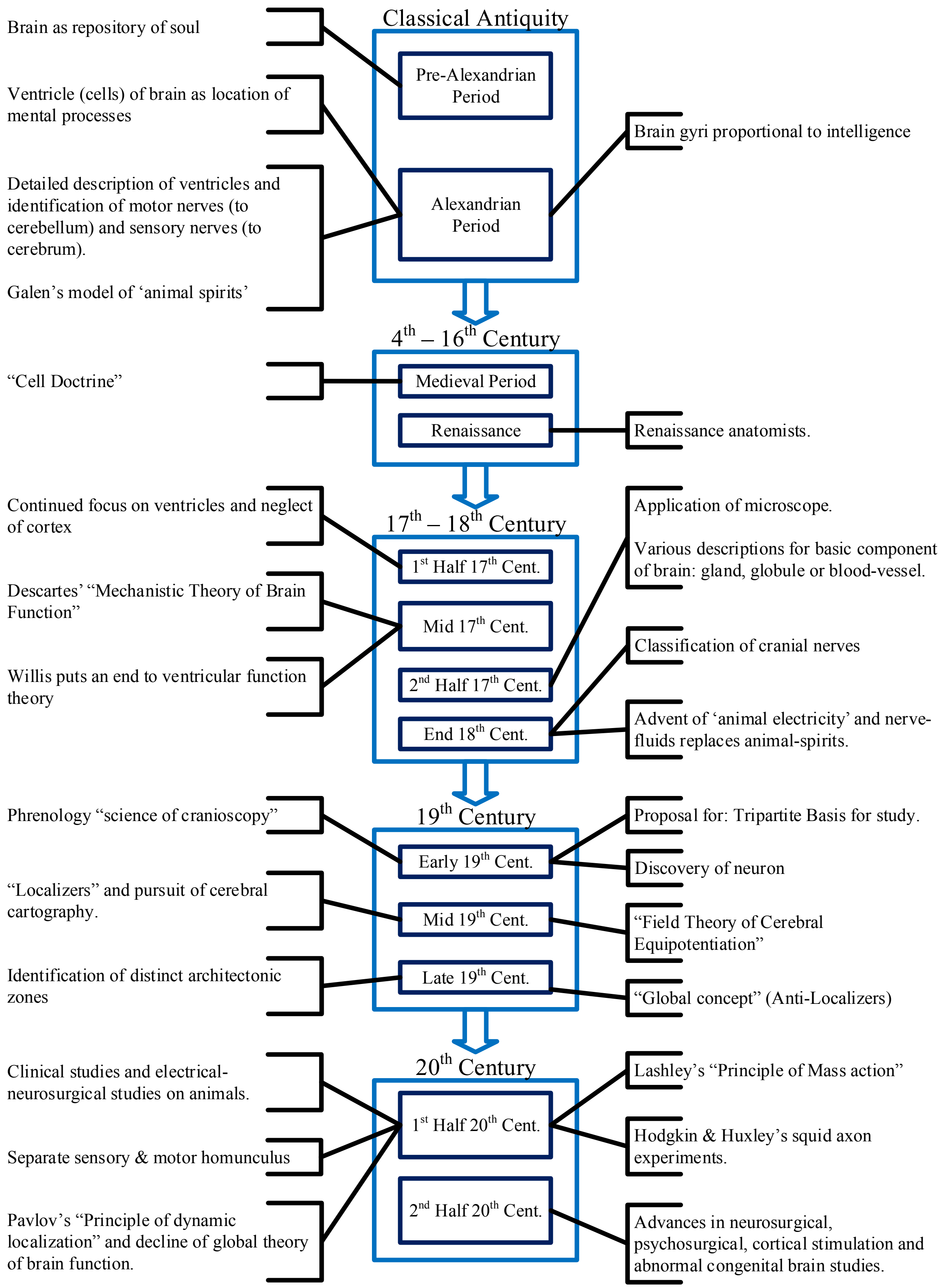


Figure 1.2. Illustration on brief history of attempts to identify and localize brain function.

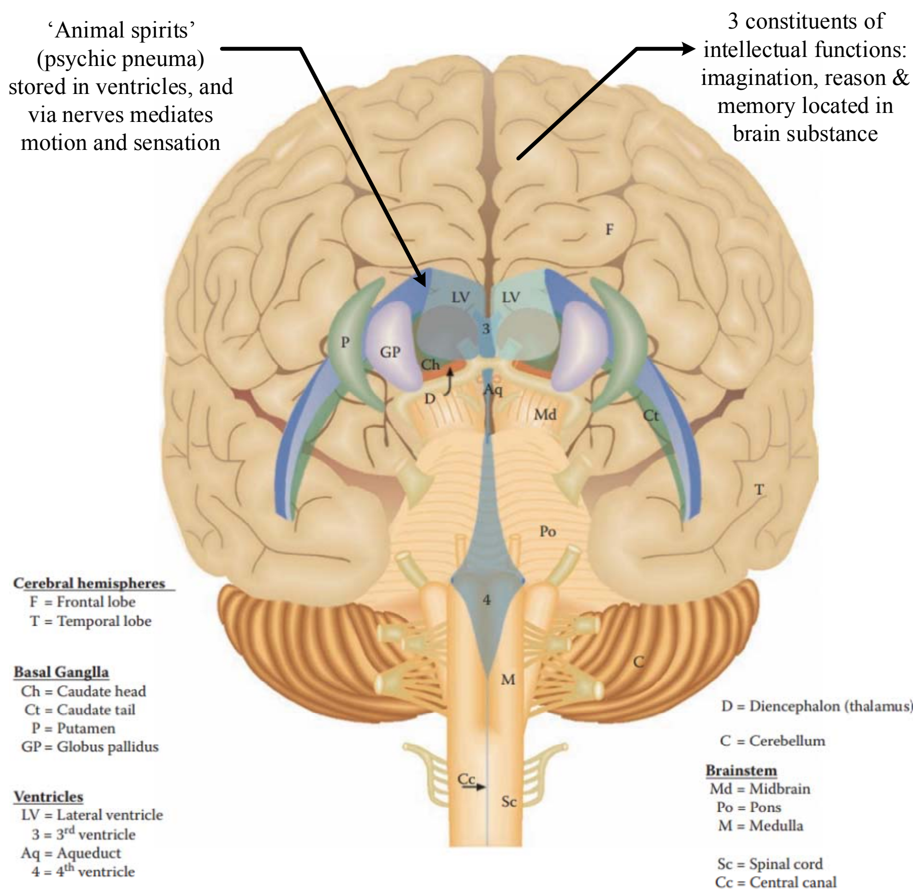


Figure 1.3. Galen's model. According to Galen, the heart produces and distributes 'vital spirits' which through internal carotid arteries and then rete-mirabile (or marvelous net, thought to be fine network of vessels at base of brain first introduced by Herophilus but not proven until 16th – 17th century that it is not a feature of human, monkey or rodent brains) enter the brain. In the brain, by process of refinement 'vital spirit' gets transformed to 'animal (animus) spirits' (psychic pneuma) which is then stored in the ventricle. It should be noted that Galen often contradicted himself. For instance, Galen also postulated that 'animal spirits' were produced in ventricles [Clarke & Dewhurst, pp.5, 1972].

The early church fathers (390 A.D, Nemesis Bishop of Emesia and 354 – 430 A.D, St. Augustine) accepted Herophilus's view of ventricles as the seat of mental processes and refined the theory into the 'cell doctrine' of brain function [Clarke & Dewhurst, pp.10, 1972] (Fig.1.4). With passing centuries the 'cell doctrine' was accepted with variations, such as the addition of dynamic element in 10th century or Avicenna's more complicated rearrangement in a five-cell scheme in the 14th century [Clarke & Dewhurst, pp.30, 1972].

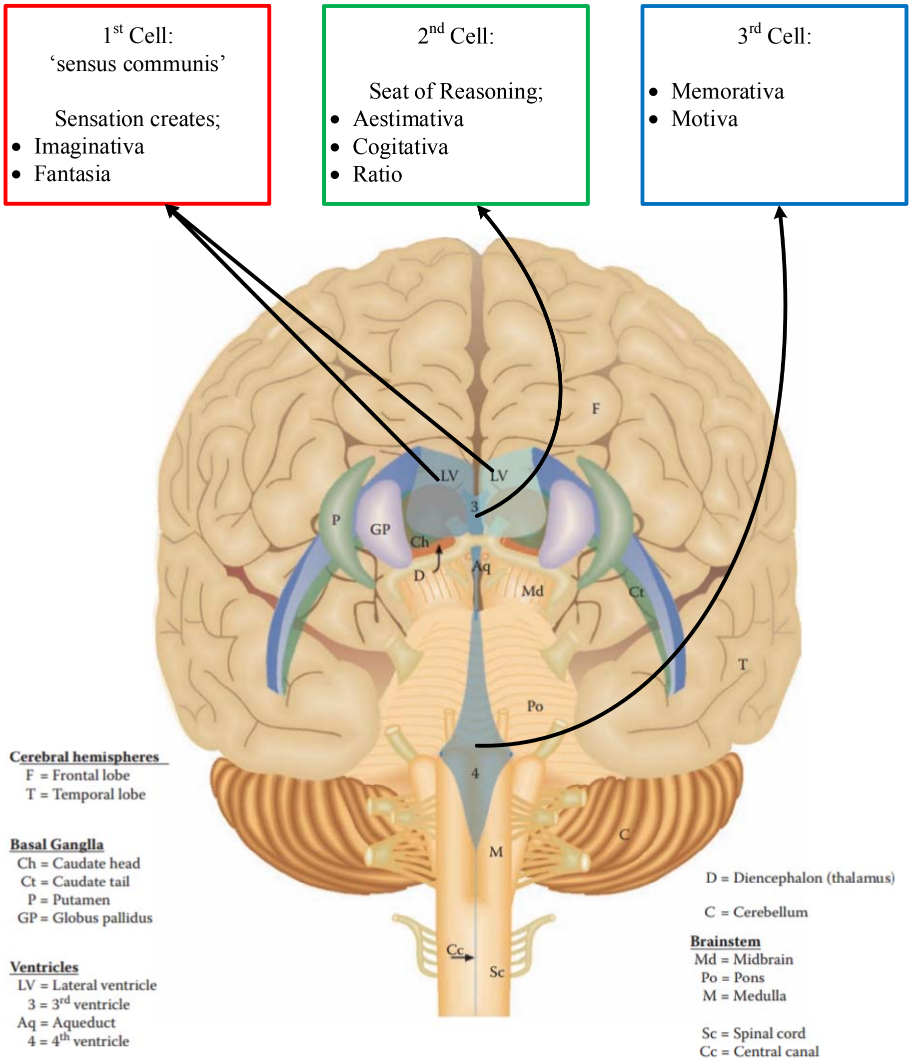


Figure 1.4. Simplest form of the medieval cell doctrine. First cell (lateral ventricles) is the seat of 'sensus communis' ('common sense'), second cell (3rd ventricle) is the seat of reasoning and third cell (4th ventricle) is the seat of memory and action. The first cell receives sensation and creates in the posterior part, imaginativa (imagination) and fantasia (image formation). Second cell is where aestimativa (judgment), cogitative (thought) and ratio (reason) takes place. Finally, third cell is the place of memorativa (memory) and motiva (motion) [Clarke & Dewhurst, pp.10, 1972].

The dynamic element added in the 10th century considered a sequence of events comparable to digestion starting from 1st cell and ending in the 3rd cell.

The renaissance anatomists (Leonardo da Vinci, Jacopo Berengario da Carpi, Andreas Vesalius, Charles Estienne, etc…) re-established original true anatomy of the Alexandrian period and the crude medieval sketches were no longer accepted. Increased understanding of form and function of ventricular system lead to doubts on the 'cell doctrine'. On the other hand, they re-introduced the concept of 'animal spirits' [Clarke & Dewhurst, pp.51, 1972].

In early 17th century there was continued focus in ventricular system over cerebrum though there was decreased support for the 'cell doctrine' [Clarke & Dewhurst, pp.68, 1972]. By mid-17th century the cerebrum was seen as the site for some special functions. Rene Descartes's 'mechanistic theory' of brain function focused on ventricular system but pineal body was considered for functional significance as the seat of soul and governing center [Descarte, 1664]. Franciscus de le Boe (Sylvius) and Thomas Willis independently proposed that the cerebral cortex has some special function. The former theorized that 'animal (psychic) spirits' are secreted from cerebral and cerebellar cortices [Sylvius, 1963]. Willis is however credited for putting an end to the ventricular function theory [Dewhurst, 1980] and for theorizing three localized mental functions.

The Danish anatomist Niolaus Steano however attacked Descarte [De Ninville, 1669] and also criticized Willis for continued adherence to tradition [Dewhurst, 1968]. Samuel Thomas Soemmerring, noted for clarity and accuracy of his illustration, was responsible for present classification of cranial nerves [Vandenhoeck, 1778] but regressed to medieval theory of ventricular function, of which Goethe and Kant were critical and challenged any attempt to localize soul [Clarke & Dewhurst, pp.85, 1972]. Steano called for a program of research which made good sense [De Ninville, 1669] but there was little or no investigation on cerebral cortex by end of 18th century and hence no notable change in brain function concepts [Clarke & Dewhurst, pp.81, 1972]. However, the application of microscope stimulated research resulting in advent of 'animal electricity' (nerve-fluid replaces animal spirits) portending revolutionary change in nerve tissue and brain function studies [Brazier, 1958].

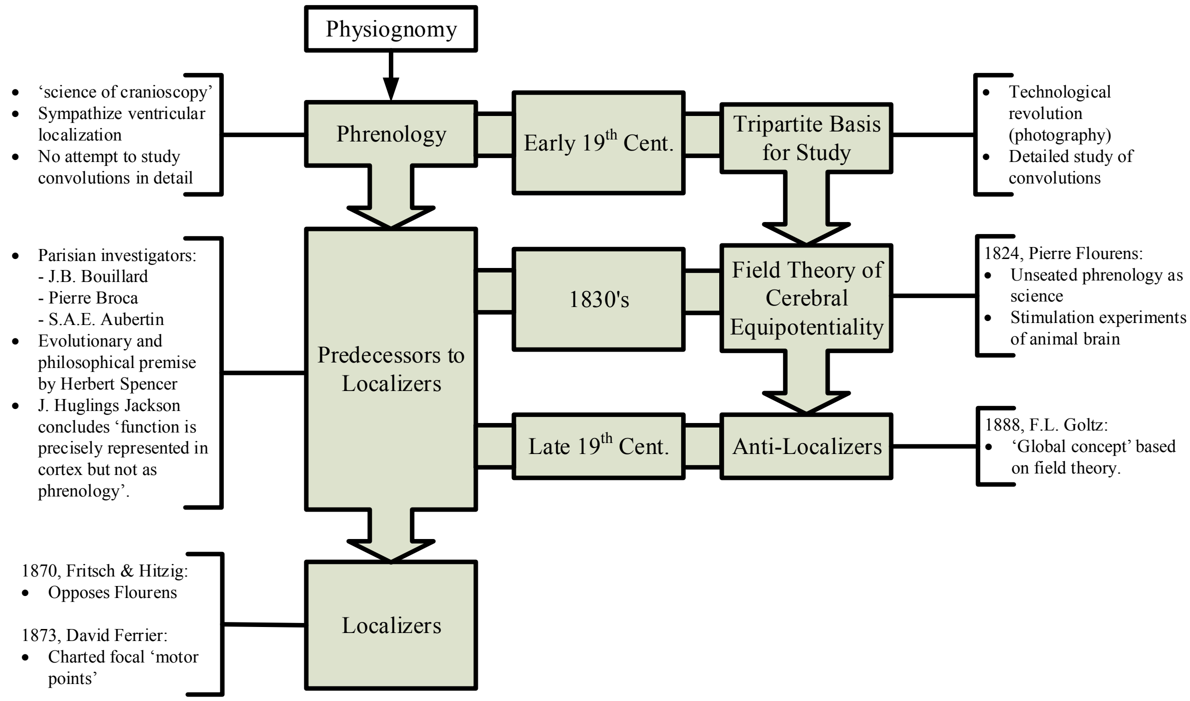


Figure 1.5. Development of investigative approaches in studying the role of cerebral cortex during the 19th century. Physiognomy is "the detailed study of cranial shape and size as a supposed indication of character and mental abilities" [Soanes & Stevenson, 2005]. During the development and increased popularity of phrenology, French and German anatomists led the tripartite study taking advantage of the contemporary technological revolution and the photographer later replaced the artist as the middle man between the anatomist and the reader [Clarke & Dewhurst, pp.106, 1972]. Flourens, using stimulation experiments, unseats phrenology and advances the "field theory of cerebral equipotentiation", which preceded the 'global concept' of Goltz.

The anatomist studied the notion of localizing specific brain function which led to the pioneers, Fritsch & Hitzig, providing experimental support. They did this by eliciting contralateral limb movement following galvanic cortical stimulation of dog brain. They also demonstrated disturbance of limb motor function following removal of the delineated cortical areas. By late 19th century the localizers were not unopposed as the anti-localizers supported 'global concept'.

The 19th century saw developments in the idea of the role of cerebral cortex in brain function that would become the roots of modern era of concepts (Fig.1.5). In early 19th century, physiognomy was transformed by Franz Joseph Gall into phrenology, the science of cranioscopy, but based on unscientifically and naively selected and interpreted evidence [Schoell, 1810]. By correlating mapped areas on cranium (physiognomy) to organ (mental faculties) numbers directly placed on gyri, phrenologists brought together two themes; localization of brain function and morphological arrangement of gyri. Later, the organ boundaries were precisely marked on the brain [Theoré, 1836], thus linking phenology to 20th century cytoarchitectonics [Clarke & Dewhurst, pp.94, 1972].

Contemporaneously, French and German anatomists subscribed to the tripartite (macroscopical, embryological and comparative anatomy) basis of brain research. The French established anatomical interpretation of disease (roots of modern medicine) concurrently developing basic medical sciences and were later joined by the surpassing Germans [Clarke & Dewhurst, pp.101, 1972]. The most renowned French anatomist was Louis Pierre Gratiolet [Broca, 1865; Grandeau, 1865]. Using comparative and evolutionary approaches, he distinguished primary from secondary gyri and defined the limits of ‘frontal’, ‘temporal’, ‘parietal’ and ‘occipital’ lobes, terms introduced by Friedrich Arnold [Höhr, 1838].

Lack of consistency and overenthusiasm among phrenologists pushed phrenology to extremes and by the 1830’s it began to lose favor. Pierre Flourens unseated the notion of functional localization in cortex and hence is credited for unseating phrenology as a science [Olmsted, 1953]. Edwin G. Boring summarizes the contribution of phrenologists as

> “the theory of Gall and Spurzheim is, however, an instance of a theory which, while essentially wrong, was just enough right to further scientific thought … away from the concept of the unsubstantial Cartesian soul to the concept of the more material nerve function … was wrong only in detail and in respect of the enthusiasm of its supporters” [Boring, 1957].

Flourens using stimulation experiments of animal brains showed that the cerebrum was inexcitable with intellectual and perceptual functions diffusely represented throughout the hemisphere although functions were located in various parts of the brain [Flourens, 1824]. This is the 'field theory of cerebral equipotentiation' [Tizard, 1959].

On the other hand, Gratiolet Parisian anatomists (Pierre Broca amongst them) focused on the problem of localizing language function as a test-case assuming that establishing existence of one function implies others would be identified in time [Soury, 1899; Moutier, 1908]. A contemporary anatomist in London, J. Hughlings Jackson [Broadbent, 1903], based on clinical – pathological data and the evolutionary and philosophical premises of Herbert Spencer [Clarke & Dewhurst, pp.497 – 498, 1968], concluded that brain function is precisely represented in cortex but not as phrenology suggested [Clarke & Dewhurst, pp.113, 1972]. This transition from phrenology to scientific investigation of specific brain function, but still considering cerebral cortex as the seat of function, was aptly summarized by Turner as,

> "the precise morphological investigations of the last few years into cerebral convolutions have led to the revival in Paris of discussions, in which the doctrine of Gall and his disciples – that the brain is not one but consists of many organs – has been supported by new arguments and the opinion has been expressed that the primary convolutions, at least, are both morphologically and physiologically distinct organs" [Turner, 1866].

During this same period, histologists increased our knowledge of the neuron [Clarke & Dewhurst, pp.67, 1968]. Building upon his predecessors, Theodore Meynert contributed to our knowledge of cortical layers of cells and fibres [Clarke & Dewhurst, pp.423 – 437, 1968]. This was followed by histologist Ramón y Cajal identifying the individual cortical constituents [Cajal, 1909].

The experimental support for the notion of local representation of function in cerebral cortex was provided by the investigators Gustav Theodor Fritsch and Eduard Hitzig [Fritsch & Hitzig, 1870]. They opposed Flourens and demonstrated excitability of cortex and located motor function. Their findings and interpretations published in 1870 began an era of progress and continues to revolutionize ideas of brain function [Clarke & Dewhurst, pp.113, 1972].

Amongst the British school of localizers, David Ferrier, pursuing Fritsch and Hitzig's view, tested and confirmed Hughlings Jackson's notion based on concepts of cortical localization for etiology of unilateral epilepsy [Ferrier, 1873]. He did this by performing animal experiments in West Riding Lunatic Asylum and charted local 'motor points', later transposing monkey brain findings to the German anatomist Alexander Ecker's human brain outline [Ranke, 1887]. However, there was little or no experimental evidence of human cortical function at that time though Ferrier's animal experiments are still valid [Bartholow, 1874].

Ferrier, along with the localization pioneers Fritsch and Hitzig, triggered a new field of neurophysiology with clinical undertones. This was followed by the new field of brain surgery [Scarff, 1940], thus inspiring physiologists to study the brain and neurosurgeons in mapping the human brain. The localizers were however not un-opposed. The other school of thought, 'global concept' led by F.L. Goltz, was based on the 'field of equipotentiation' [von Bonin, pp.118 – 158, 1960]. According to the 'anti-localizers' precise localization of cortical functions was impossible because of the massive inter-neural connections and hence they opposed the popular pursuit of cerebral cartography [Clarke & Dewhurst, pp.115, 1972]. The arguments between the two schools of thoughts became public in the encounter between

Ferrier and Goltz at the International Medical Congress held in London 1881 [Wilkins, pp.119 – 129, 1965].

Amongst the anatomist supporting localization, Meynert was the first to relate regional cortical structural differences to functions [Meynert, 1890]. By late 19th – early 20th century architectonics, the mircroscopical study of appearance of cells and fibers was used to identify morphologically distinct areas and argued that the direct correlation with function can be found [Fulton, 1949]. The two pioneers were Walter Campbell and Korbinian Brodmann, both concerned with advancement of knowledge of function and pathological manifestations [Haymaker & Schiller, 1970]. Compared to his German contemporaries Campbell (Australian) was more conservative in identifying distinct architectonic zones and viewed that histology must follow physiologists and clinicians to provide more accurate demarcations of areas corresponding to cortical function [Campbell, 1905]. Brodmann, more concerned with comparative studies, identified 52 areas grouped into 11 histological regions [Broadman, 1909] and helped bring order into the confused state of knowledge [Clarke & Dewhurst, pp.121, 1972]. Oskar and Cécile Vogt with their cyto-architectonic studies of primates and man with electrical stimulation studies dominated the field of cortical localization in early 20th century [Haymaker, 1951]. Otfrid Foerster, a neurosurgeon with interest in localization and excision of epileptogenic foci, advanced the understanding of human cortical physiology and helped dissipate Vogt's confusing results [Zülch, 1969].

Non-localizers criticized architectonics. K.S. Lashley and G. Clark were the most vociferous amongst them [Lashley & Clark, 1946]. Using rat maze studies [Lashley, 1929] Lashley supported the 'theory of equipotentiation' and considered the principle of mass action as applied to intelligence. But with accumulating clinical studies he modified his view to accommodate regional subdivisions of verbal and non-verbal learning abilities [Lashley, 1938]. The concepts of 'functional pleuripotentialism' and 'graded localization of functions' are considered the basis for Pavlov's principle of 'dynamic localization' [Luria, pp.27, 1966]. I.N. Filiminov, from his study of dolphin olfactory structures developed 'functional pleuripotentialism' which views that no part of the cortex is solely responsible for single nervous function but under certain conditions each part of brain may perform other functions [Filiminov, 1961]. H. Nakahama identified four somatotopical regions, each concerned with sensory and motor functions [Nakahama, 1961], thus providing supportive evidence for theory of multiple cortical function.

Thought acceptance of the 'global concept' began to decline [McFie, 1972], some psychologist [Vernon, 1950; Bruner, 1964] still supported Lashley's view and claimed that environmental factors mostly determine intelligence, rejecting any possibility of correlation with specific brain areas. Others such as Chapman and Wolff tried to reconcile the two theories with their not antagonistic but complementary theory. However, Wilder Penfield, initially working with Foerster, continued his work at Montreal with his colleagues making one of the most important contributions to knowledge of cortical localization of functions [Clarke & Dewhurst, pp.127, 1972]. Using electrical stimulation studies of conscious patients Penfield et al. summarized the results of illustrating order and comparative extent of cortical representation of elements in sensory and motor sequence in sensory and motor homunculus [Penfield, 1937; Penfield, 1957]. In attempts to understand brain function contemporary clinical and experimental psychologists continue to investigate for cortical localization but with techniques having better temporal and spatial resolutions [Gazzaniga et al., 2006].

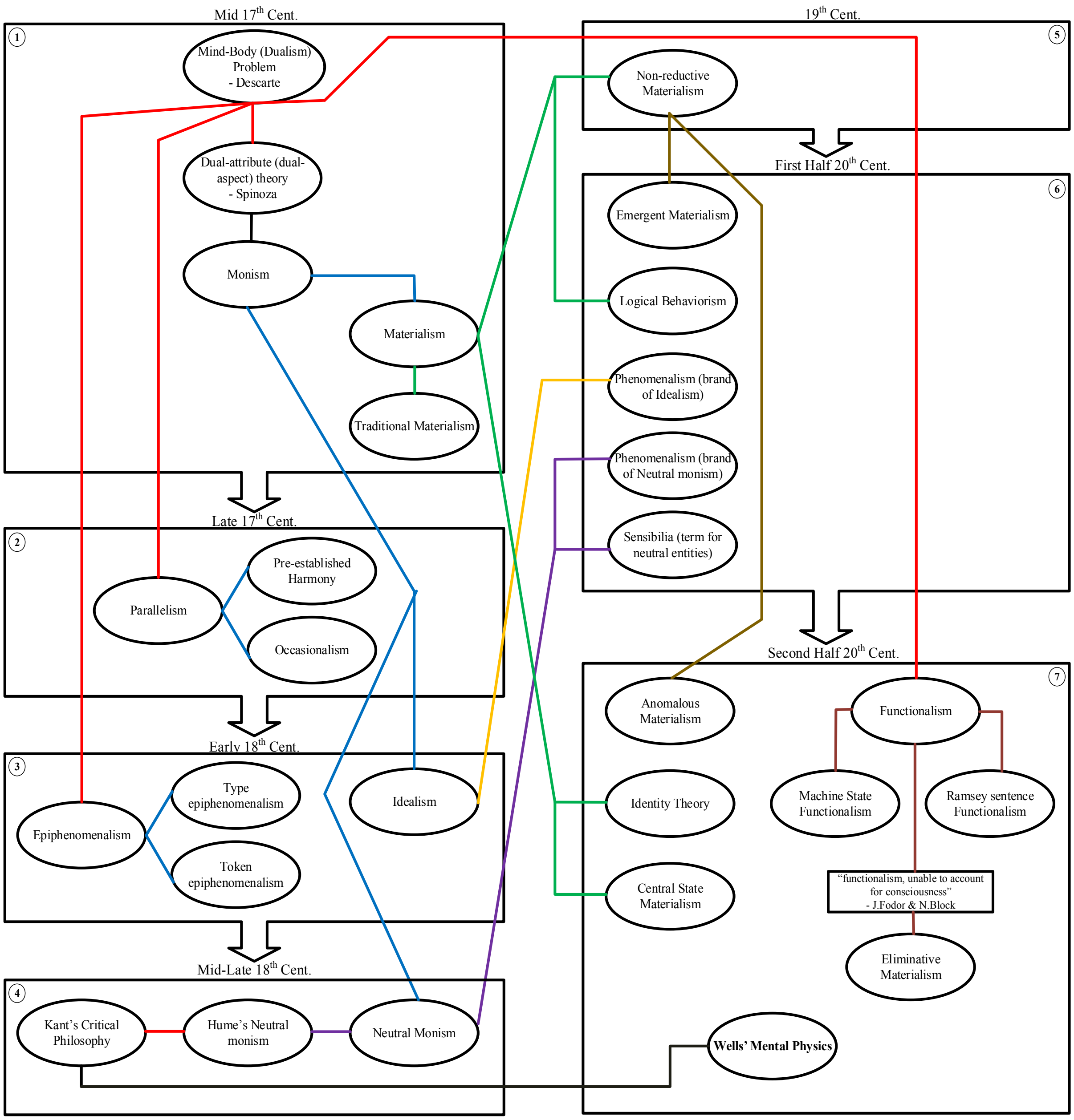


Figure 1.6. Summarized illustration of concepts of philosophy of mind. Blocks (labelled 1 – 7) represent respective period starting from mid-17th to 20th century. Blocks on the left span from mid-17th to 18th century. Ovals represent concepts with lines joining them. Black lines represent evolved or transformation in concept. Red lines represent argument against the preceding concept. Other colored lines represent brands of their respective preceding concept.

Philosophy of Mind and Psychology

The preceding paragraphs indicates inter-linkages between ideas generated from 'laboratory or investigative' works (anatomy, biology, psychology, etcetera) with philosophy. However to avoid confusion an overview of philosophical ideas concerning 'how brains work?' is discussed separately, as illustrated in figure 1.6.

René Descartes is considered father of the modern mind-body problem though the doctrine of distinct soul from body was discussed throughout history [Audi, 1995]. *Cartesian dualism* views minds as substances that are not extended in space and hence are distinct from any physical substance. The mid-17th century Descartes's dualistic view resulted in generation of other philosophical ideas through the centuries, mostly to argue against dualism (Fig.1.6).

Baruch Spinoza, rejecting Descartes's bifurcation of reality into mental and physical substances, held a '*dual-attribute*' of '*dual-aspect*' theory [Audi, 1995]. This view refers to God as the single substance with distinct modes, mental and physical. However, some philosophers opted for the view that 'all reality is really of one kind', also called *monism* [Audi, 1995]. Thomas Hobbes, a contemporary of Descartes, considered 'everything is material or physical'. This is *traditional materialism*, a brand of monism [Audi, 1995].

Contemporaries of Spinoza, particularly Nicolas Malebranche and Gottfried Wilhelm Leibniz, also arguing against dualism considered *parallelism*, 'mental and physical realms run in parallel' (Fig.1.6, block2) [Audi, 1995]. Thus due to God's creation the types of mental phenomenon co-occur with certain types of physical phenomenon but never involve causal interaction. Leibniz's brand or *pre-established harmony* viewed that co-occurrence is only possible in world actualized by God. Similarly Malebranche's brand or *occasionalism* viewed that non-divine phenomena never cause anything and only God's activities cause things to happen.

To solve the mind-body problem while accepting the dualistic view Julien Offray de La Mettrie considered a one-way psychophysical action with physical state causing mental state but not the other way. Thus what we perceive must cause us to undergo sensual experience. This is *epiphenomenalism* [Audi, 1995]. 19^{th} and 20^{th} century materialism raised the issue of whether any state as causes ever fall under mental types. This resulted in two theses; *type epiphenomenalism*, 'no state can cause anything in virtue of falling under a mental type' and *token epiphenomenalism*, 'no mental state can cause anything' [Audi, 1995].

Apart from Hobbes's traditional materialism another brand of monism is *idealism*. George Berkeley's idealism viewed that 'everything (mental and physical phenomenon) is mental, perceptions in God's mind' [Audi, 1995]. Georg Wilhelm Fiedrich Hegel's brand modified this to 'everything is part of the World Spirit' [Audi, 1995]. *Neutral monism*, also a brand of monism considers 'all reality is ultimately of one kind neither mental nor physical'. According to David Hume, 'mental and physical substances are really just bundles of neutral entities' [Audi, 1995]. However in the 20^{th} century Bertrand Russells's version or *sensibilia* viewed 'mind and physical objects as logical constructs out of sensibilia' [Audi, 1995].

Philosophers acknowledged two major problems with monism: characterization of fundamental entities and explaining how they make up non-fundamental entities [Audi, 1995]. With the rebirth of atomic theory and quantum mechanics proponents of materialism believed in its success where idealism and neutral monism failed. In the late 19^{th} century *non-reductive materialism* considered that the 'every substance either is or is wholly made up of physical particles'. Thus the well-functioning brain is the material seat of mental capacities and token mental states are token neurophysiological states. Charles Dunbar Broad's brand or *emergent materialism* viewed 'mental capacities, properties, etcetera emerge from, and thus do not reduce to physical capacities, properties, etcetera' [Audi, 1995; Gustavsson, 2010]. In the second half of 20th century Donald Davidson's brand or *anomalous monism* is a combination of the thesis that 'there are no strict psychological or psychophysical cause' and his 'irreducibility thesis' [Audi, 1995; Malpas, 2013]. According to irreducibility thesis 'every event token is physical but intentional mental predicates and concepts do not reduce to physical predicates or concepts'.

To solve the dualist point that 'one can understand ordinary psychological vocabulary like belief, desire, pain, etcetera and yet know nothing at all about physical states and event in brain' the materialistic doctrine was *logical behaviorism* [Audi, 1995]. They viewed that 'mental states are not internal states with causal effects but are phenomenon that is shorthand for actual or potential overt bodily behavior'. This was much discussed from around 1930's to early 60's with their proponents like Gilbert Ryle and Rudolf Carnap ridiculing Cartesianism's view as 'ghost in the machine (body)' [Ryle, 1949].

Herbert Feigl, a materialist, claimed that mental states are brain states and terms differ in meaning but scientific investigations reveals same referents (morning-star & evening-star differ in meaning but both refer to Venus) [Feigl, 1967]. J.J.C. Smart and U.T. Place, defending this view, introduced the identify theory which considers that 'sensations are identical with brain processes, knowable only by empirical findings' [Audi, 1995]. Another contemporary materialist, David Armstrong supported *central state materialism* which took 'mental states as contingently identical with states of brain apt to produce certain range of behavior' [Audi, 1995; Jackson, 2006]. Thus, compared to logical behaviorists, central state materialists hold mental states as actual internal states with causal effects without implying translation of the mental to behavior, but unlike Cartesianism hold psychophysical interactions are just physical causal interaction [Audi, 1995].

A.J. Ayer supported *phenomenalism*, a view that 'all empirical statements are synonymous with statements solely about phenomenal appearance' [Macdonald, 2010]. Thus if the phenomenal appearance is claimed to be neither mental nor physical it is a brand of idealism and brand of neutral monism if claimed mental [Audi, 1995].

Another argument against dualism is *functionalism* which considers 'specific mental types are types that play a certain causal role'. In the 1960's Hilary Putnam proposed a brand now called *machine state functionalism*. This is a view that 'mental states are types of Turing machine table states' [Putnam, 1995]. Contemporaneously David Lewis proposed *Ramsey sentence functionalism*, a view that 'conjunction of commonsense psychological platitudes to formulate Ramsey sentence defines all mental predicates in platitudes in physical and topic-neutral terms' [Audi, 1995; Weatherson, 2010]. Jerry Fodor and Ned Block raised the problem of functionalism not being able to account for consciousness [Audi, 1995; Block & Fodor, 1972]. Patricia Churchland responded with *eliminative materialism*, a view that 'denies any mental phenomenon, i.e., eliminativist (there is no such thing)' [Audi, 1995; Warburton, 2010]. The philosophy espoused in this present research is mental physics, developed by Wells based on Kantian epistemological metaphysics [Wells, 2009]. This will be separately described later.

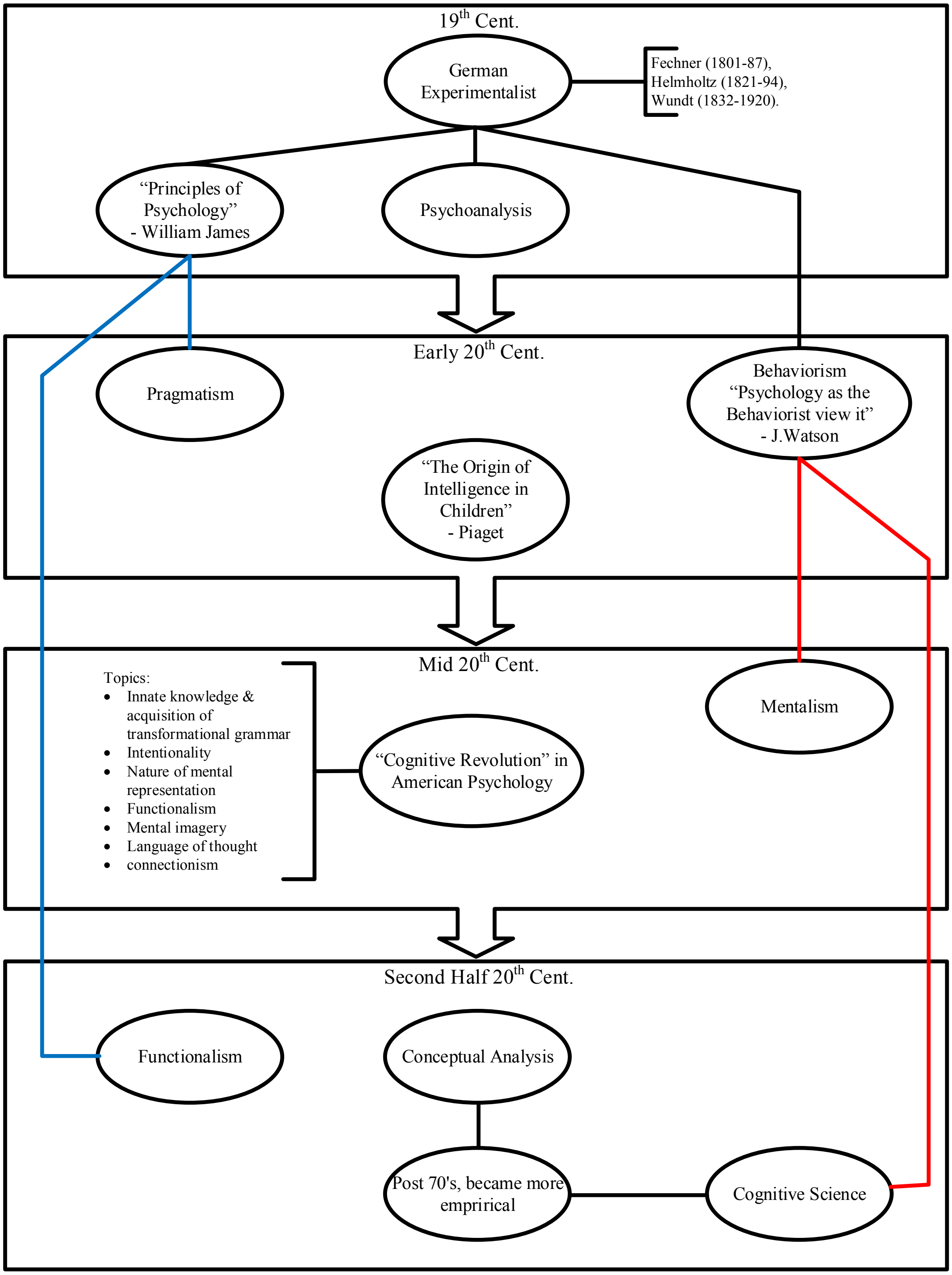


Figure 1.7. Summarized illustration of concepts of philosophy of psychology. Blocks represent respective period starting from 19th to 20th century. Ovals represent concepts with lines joining them. Black lines represent evolved or transformation in concept. Red lines represent argument against the preceding concept while blue lines represent brands of respective preceding concept.

Psychology as the Rebers describe it is philosophers' and scientists' creation "to fulfil the need to understand the minds and behaviors of various organisms, from the most primitive to the most complex" [Reber & Reber, 2001]. Following German experimentalists in the 19th century, psychology began to separate from philosophy (Fig.1.7) [Audi, 1995]. Sigmund Freud, an Austrian neurologist now recognized as the father of psychoanalysis, proposed arguments that continue to contribute to the human understanding of 'how brains work?' [Kandel, 1999]. William James, identified as the father of American psychology, was familiar with the works of the German experimentalist and wrote the classic books 'The Principles of Psychology' (2 volumes) in 1890 [James, 1890]. James's work does not have a structured theory but the adjectival form to his empirical and philosophical underpinnings are *pragmatism* and *functionalism* [Reber & Reber, 2001]. Pragmatism views that 'meanings and truths of propositions are taken as equivalent to practical outcomes' [Reber & Reber, 2001].

James Watson in his 1913 paper 'Psychology as the Behaviorist Views It' supported autonomy of psychology from philosophy and criticized psychologist and experimentalist for relying on introspective methods and making consciousness the discipline's subject matter [Watson, 1913]. Behaviorism is the approach to psychology as a 'purely objective experimental branch of natural science with the theoretical goal to predict and control behavior discarding all references to consciousness.' However the anti-behaviorist argued that if the attempt to explain behavior is legitimate, can the above goal be met without appealing to mentalism ('any theory couched in terms of mental events and processes')? Ironically behaviorists like B.F. Skinner responded with no empirical evidence but gave philosophical arguments to ban mentalistic causes [Audi, 1995].

Jean Piaget published series of classic books and is best known for describing the four development stages from studying psychological development in children [see citations]. He called his theory of cognitive development genetic epistemology. Between around 1945 and 1975 'conceptual analysis' (process of breaking down concepts into simple parts displaying its logical structure) dominates American and English philosophy of psychology peaking with the works of logical positivism [Audi, 1995; Blackburn, 2005]. They took the position that 'philosophy is essentially an a priori discipline' but rarely cited empirical studies. After 1950 the 'Cognitive Revolution' in American psychology (an intellectual movement) highlighted set of topics beyond just mind-body problem and behaviorism and cognitive psychology replaced the latter [Leahey, 1992; Mandler, 2002].

By 1970's behaviorism began to decline and conceptual analysis became more empirical. Coincidentally 'cognitive science' emerged as a new discipline or more accurately 'a cluster of disciplines like cognitive psychology, epistemology, linguistics, computer sciences, artificial intelligence, mathematics and neuropsychology that study the human mind'.

In summary the above overview of philosophical arguments indicate that there is no single argument unanimously supported amongst philosophers. This has led to an aversive reactions from people in disciplines that have separated from philosophy. Using the problem of ambiguity amongst psychologists on the semantic root of "instinct" as an example Wells points out,

> "Unfortunately, almost all the early twentieth century psychologists were eager disciples of positivism and embraced with enthusiasm its *ex cathedra* dogma of *ignórance* that: (1) science could learn nothing from philosophy; and (2) physics was "the queen of all the sciences". Crippled by their ontology-centered prejudices and baselessly subjective pseudo-metaphysics, they managed to turn

> the idea of "instinct" into such a mishmash of vague and Platonic notions that the subsequent usages of the term actually became counterproductive. Then then, like Aesop's fox, declared the whole idea to be sour grapes and, like medieval inquisitors prosecuting heresy, burned the idea at the stake and scattered its ashes on unhallowed grounds. Psychology has never recovered from the trauma of this historic episode and preserves the transcripts of those ecclesiastical court proceedings in the prosecution of "instinct" in school doctrines today." [Wells, 2011b].

Past and current works on determining equivalence or equivalence like problems

Psychological Works

Following behavioral findings and interpretation from experiments performed by linguists, some behaviorist in the late 20th century studied equivalence. Some of the behaviorists viewed that equivalence relation testing for conditional relation between stimuli such as auditory and visual stimuli provides a behavioral basis for studying language comprehension and production [Bush, 1993]. They defined equivalence class as classes containing finite number of stimuli (N) bearing no overt perceptual similarity but becoming related by baseline training N – 1 conditional discriminations among N stimuli [Sidman, & Tailby 1982; Fields & Verhave, 1987; Sidman, 1994; Fields & Reeve, 2001]. They claimed that, after baseline training, tests evoked emergence of untrained relations, which is an equivalence relation when a particular stimulus evokes selection of a different stimuli which are functionally interchangeable [Fields & Verhave, 1987; Sidman, 1994]. Figure 1.8a illustrates this.

In other words, manipulation of reinforcements can establish arbitrary relation between a given response and stimuli [Lazar, 1977]. They defined such stimuli to be functionally equivalent [Goldiamond, 1962]. Some behaviorist prescribed to "mediation theory" which is an approach to the study of learning which assumes that some event(s) intervenes between response and stimulus. Explication of this intervening process is required to explain behaviors [Reber & Reber, 2001]. They claimed that "mediation-transfer" experiments showed that stimuli can become functionally equivalent by means other than training [Jenkins, 1963].

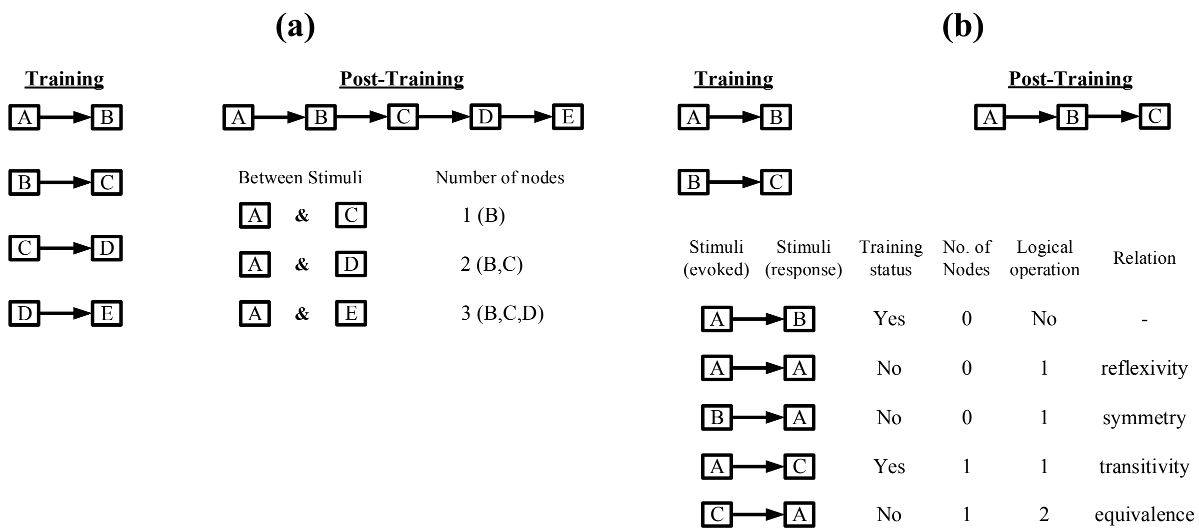


Figure 1.8. Illustration of evoking equivalence and other relations after training. In accordance to their definitions, the structure of equivalence class is defined by four atemporal variables; class size, number of nodes, training directionality and nodal density [Fields & Verhave, 1987]. The variables either independently or in conjunction with each other influence the level of relatedness.

(a) Training for a class size of 6 such that each pair is trained by evoking a response (eg. B) for a given stimulus (eg. A). After training, a stimulus (say, A) can evoke response that was not trained (say, C). The figure also shows how the number of nodes between stimulus and response are counted. Beside the number the elements inside the bracket are the respective nodes.

(b) Training for class size, 3. The table shows some of the possibilities after the training. A→B is already trained and hence no logical operation is considered to be performed. For A→A & B→A the evoked stimuli must connect to A. Hence one logical operation is performed for reflexivity & symmetry. One logical operation is also performed for transitivity because A must connect to B but after that connection of B to C is already trained. Finally, C→A is equivalence relation with two logical operations.

If simple as defined with regard to lesser logical operations then, transitivity is a simpler relation than equivalence relation. Some view that, because of this, transitivity relation will be preferred to more complex equivalence relation [Doran & Fields, 2012].

There are at least two schools of thought. One view is that members of an equivalence class are equally related to each other [Doran & Fields, 2012]. There are numerous claims of support for this view [Barnes et. al., 1995; Barnes Keenan, 1993; Fields et. al., 1993; Rehfeldt & Hayes; 1998; Saunders et. al., 1988; Sidman & Tailby, 1982; Sidman et. al., 1989].

The challenging view is that different pairs of stimuli can have different levels of relatedness under some conditions [Doran & Fields, 2012]. Fields and Doran claim to support this by demonstrating participants who show exclusive preference for transitive relation over equivalence relation. They concluded that members of equivalence classes are differentially related to each other based on relation type and strength of relation is greater for simple relation [Doran & Fields, 2012] (Fig.1.8b).

Therefore, regardless of conflicting views among behaviorists, their definition of equivalence class assumes that a process or processes can be trained to yield equivalence relations. However they do not consider why or how the relations are equivalent in terms of features of the object nor do they consider the situation or context. This is because they assume an equivalence relation already exists and hence are ontology centered. Thus they are unable to explain the source for the claimed equivalences.

Jean Piaget also worked on problems related to equivalence relations as a part of logical thinking in children. In a series of experiments in 59 children from ages 5 years to adolescents he demonstrated three-stages of generating equivalence classes [Piaget et. al., ch.2, 1977] (Fig.1.9). His study was therefore on the three development stages: pre-operational, concrete operations and formal operations stage.

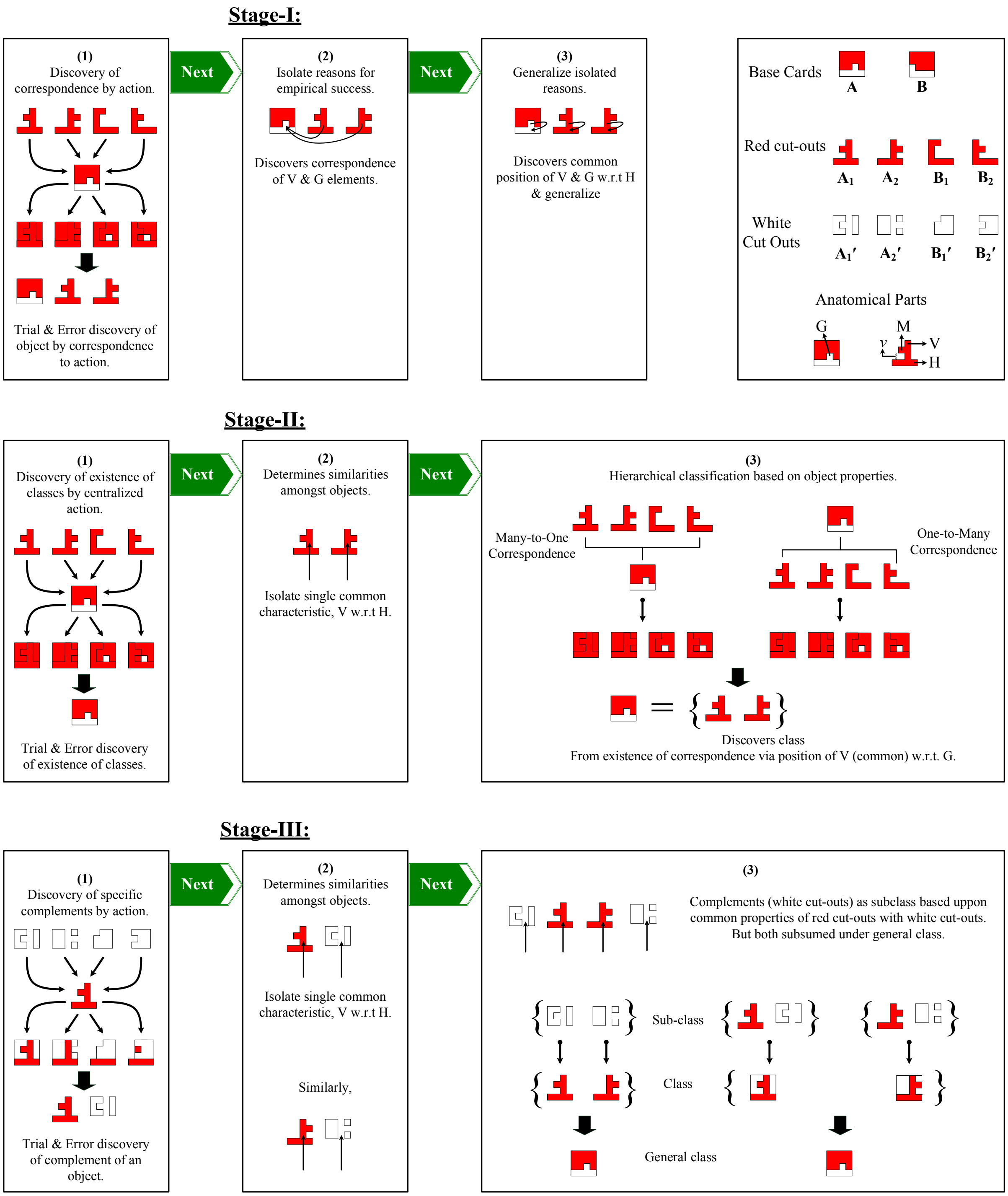


Figure 1.9. Piaget's 3-stages of constructing equivalence classes as observed in children at concrete and formal operational stages. Each stage has three sub-stages. The top far-right box shows the base cards and the cut-outs used in the description of each stages.

Stage-I is the stage of exploration and exclusive 'suitability' [Piaget et. al., ch.2, 1977] (Fig.1.9 Top). He uses the suitability for functional dependence linking the object with action. For instance, a child places a triangle in place of a square as roof of house. This stage may be subdivided into 3 sub-stages which is a passage from a given action to detailed surjective correspondence. First a functional correspondence (i.e., application) is determined by trial and error realized without any anticipatory inspection. This is followed by isolation of some features (reasons) of the object for its empirical success. There is still no anticipation or generalization of the results for subsequent attempts. Finally common features of the object are discovered and immediately generalized.

Stage-II is the stage of successive forms of equivalence classes where similarities between objects are found based upon properties of the objects (Fig.1.9 Middle). There are again 3 sub-stages passing from a functional (action) scheme to operational groupings. First existence of equivalence classes are determined by trial and error. Thus, in contrast to objective correspondence, suitability is determined relative to action. This is followed by search and isolation of a single common characteristic among objects based on appearance of objective correspondence, thus taking into account differentiated characteristics and hence determining similarities. This frees equivalence class from general scheme of action in favor of established direct relation between objects. Finally, hierarchical classification is made by reversible movements of many-to-one and one-to-many. This is done by referring to characteristic positions of the object.

Stage-III is the stage of class union operation with object complements (Fig.1.9 Bottom). Piaget argues that the problem of passing from single classification to operational grouping of classification arises because equivalence classes by simple unions do not divide themselves. To impose this new operations must intervene [Piaget et. al., ch.2, 1977]. First, complements are identified by trial and error without understanding the general relation of complementarity. This is followed by isolation of feature (or absence of feature) of the complement with regards to the object. Finally, complements with common properties corresponding to the objects form subclasses for the hierarchical construction.

Some psychologist have also researched on equivalence classes based on Piaget's INRC group of action (identity, negation, reciprocation, correlation) of formal operation stage [Inhelder & Piaget, 1958; Common, 1993]. Using the scenario of doctor-patient relationship Commons researched equivalence classes on adults with the hypothesis that classes can be arranged according to order of hierarchical complexity [Common, 1993].

Therefore, though Piaget's explanation is more rigorous than the behaviorists it still assumes that some scheme of actions exists which forms the basis for grouping operations. It also fails to explain how the common features of the objects are isolated or discovered.

This project does not implement any of the above views on equivalence. Our approach to this problem is based on Kantian metaphysics and hence is epistemological. We claim that apart from Kantian's critical acroams, the process of comparison can be explained by mental physics (application of Kant's metaphysics to phenomenon of mind study) which does not make presupposition about the source of knowledge of comparison.

<u>Mathematical Theories</u>

19th century psychophysicists studied intensity of sensations as mathematical laws to predict stimulus magnitude versus sensory discrimination. The pioneers were Weber, Fechner, Helmolz and von Frey [Kendel, 2000]. Weber in 1834 demonstrated that sensitivity depends on absolute strength of stimuli. He is best known for Weber's law, which states that the 'just noticeable difference' between the stimuli is proportional to the strength of the reference stimuli. Fechner in 1860 extended this law using logarithmic function. According to this law, intensity of sensation is proportional to the logarithmic function of the reference stimuli. In 1953 Stanley Stevens modified Fechner's version by using a power function in place of the logarithmic function (Fig.1.10).

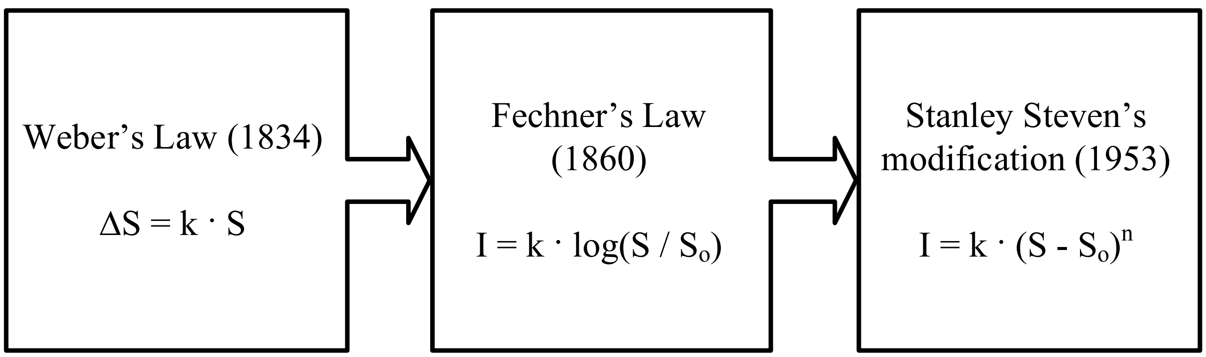


Figure 1.10. Intensity of sensation as mathematical laws. In general, S is the strength of the reference stimulus being compared against, k is a constant and I is the intensity of sensation.

ΔS in Weber's law is the minimum difference between S and another stimulus. This is often called the 'just noticeable difference' or difference limen.

S_o in Fechner's law is the threshold amplitude of the stimulus.

In the power function form of Fechner's law, if n=1 the intensity of sensation is linear. Sensory experience of hand pressure is linear [Kandel, 2000].

In the early 20th century, Nicolas Rashevsky, Herbert D. Landahl and Alston S. Householder pioneered mathematical biophysics [Rashevsky, 1938; Householder & Landahl, 1945]. Their work ranged from study of cellular diffusion and metabolism to excitation-conduction in peripheral nerves and organization in central nervous system (CNS).

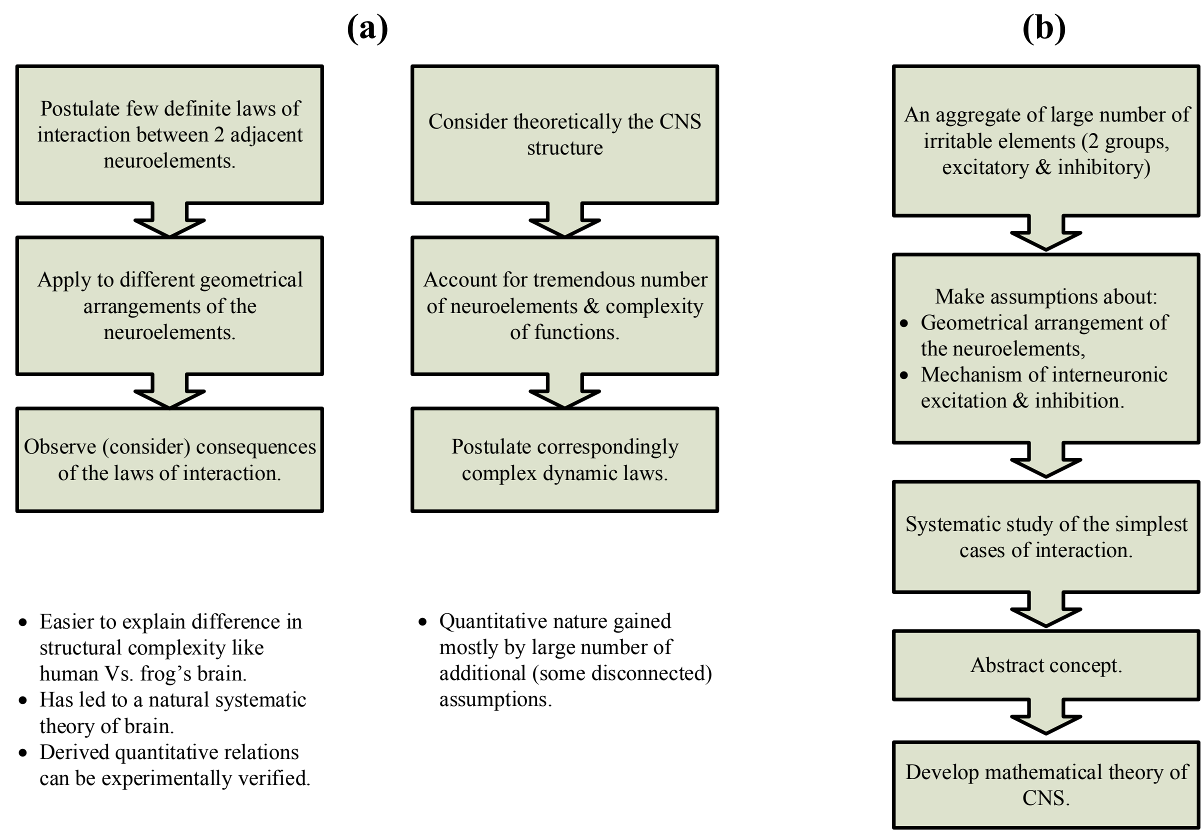


Figure 1.11. Rashevsky's approach to a systematic mathematical study of central nervous system (CNS) functions. (a) Shows the two possible approaches with commentaries about them given in bullet point underneath them respectively. (b) Elaboration of the left method shown in (a).

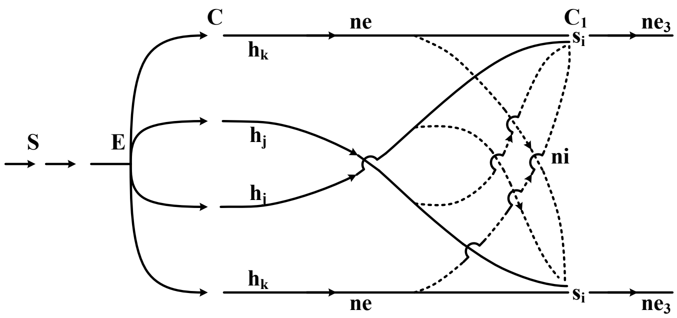
In his book Rashevsky states that a systematic abstract mathematical study of functions of the CNS can be done from two approaches [Rashevsky, pp.355 – 356, 1938]. The fundamental idea of the first method is to postulate few definite laws of interaction between two adjacent neuroelements and then consider their consequences when applied to various geometrical arrangements (Fig.1.11a, left). The second method is to consider theoretical structure of the enormously complex CNS and postulate corresponding complex dynamical laws of interaction between individual elements (Fig.1.11a, right). Rashevsky prefers the first of the two approaches because additional assumptions are needed to gain any quantitative nature from the second method while the first method reduces complexity of CNS functions to complexity of its structure yet keeping fundamental dynamic process as simple as possible.

On analyzing the abstract concepts, in the first edition of the book Rashevsky's main interest was on proving that complicated phenomenon can be systematically designed by simple systems of postulates. Householder and Landahl elaborated that though the postulated equation may not seemingly agree with experimental evidences on interaction of two neurons Rashevsky's prescribed approach (Fig.1.11b) makes good agreement with actual observations. Thus, postulated equations are more than mere mathematical assumptions and in second edition of his book Rashevsky was not much concerned with direct relation of the fundamental postulates to actual observations [Rashevsky, 1938].

Based on the fundamental postulate by Rashevsky some geometrical arrangements (neural network) can discriminate two stimuli (Fig.1.12a). With such arrangements, an intensity S yields an excitation of a definite group of connections s_i (synapse) while a different S′ excites a different group of s_i'. In Rashevsky's version (Fig.1.12a) he showed the above property by mathematical arguments and concluded that 'each intensity may be considered as a different stimulus pattern' [Rashevsky, ch.33, 1938]. A similar property but with different arrangement was discussed by Householder [Householder, 1939] (Fig.1.12b). Using Weber's ratio as a measure of just-determinable difference of total intensity it was shown that the theory compared well with experimental findings of intensity-discrimination at varying intensity levels of visual, auditory and tactile sensations [Householder & Landahl, ch.9, 1945].

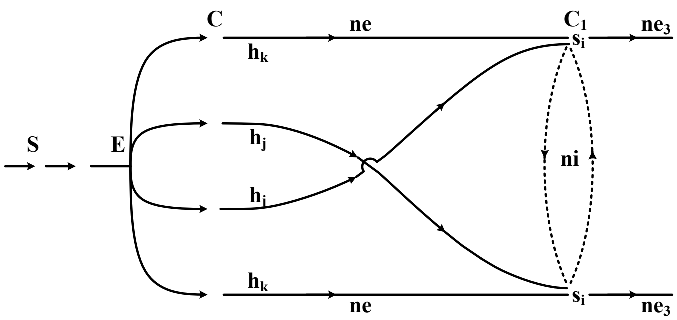
Absolute discrimination between two stimuli.

(a)



- A sensory (or chain of) pathway receives sensory stimulus of intensity S.
- The sensory pathway divides into several branches having intensity E which is same as the original pathway.
- Each branch connects to neuroelement, ne with threshold, h.
- The branches arrive at a particular region, C_1 such that:
  has same neuroelement but of higher order, ne_3,
  all ne_3 have the same threshold, h.
- Group s_i of connections in C_1 receives:
  excitation from each ne with same h_k,
  Inhibition from ni with h_j ($\neq h_k$, excited by ne).

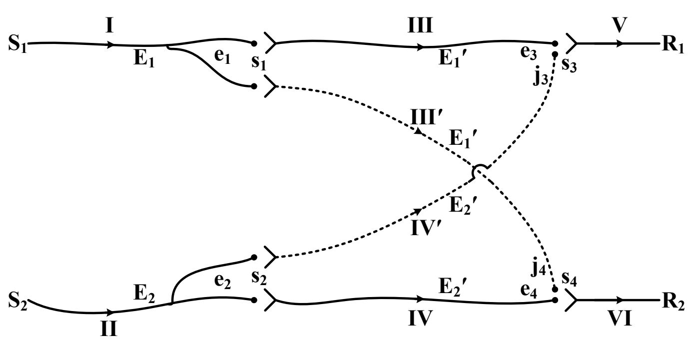
(b)



- This arrangement tackles the same problem as above.
- Arrangement uses most of the above postulates and assumptions.
- Inhibition unlike above is from only one branch.
- This illustrates that different arrangements can produce similar functions.

(c)

When absolute discrimination is difficult (2 stimuli intensities very close to each other) but differences exist



- A stimulus intensity S_1 is received by pathway I resulting in corresponding intensity E_1.
- The pathway is then connected to pathways, excitatory III & inhibitory III′ via synaptic connection s_1 having threshold h.
- The synapse s_1 has an amount e_1 (monotonically increasing function of S_1).
- When $e_1 > h$, III & III′ pathways with intensity E_1' ends at synapses s_3 (with e_3) and s_4 (with j_3) respectively having threshold h′.
- When $(e_3 - j_4) > h'$, pathway V produces response R_1.

Figure 1.12. Geometrical arrangements proposed by Rashevsky, Householder & Landahl for discriminating stimuli. (a) & (b) discriminates by receiving stimuli one at a time. (c) discriminates by receiving two stimuli simultaneously. This was introduced for problems when the difference between the stimuli is small such that (a) and (b) failed.

The above arrangement discriminated stimuli when they were presented independently. But the question then arises when two different stimuli intensities are very close to each other such that absolute discrimination become difficult. For such problems Landahl introduced a geometrical arrangement (Fig1.12c) such that it receives two stimuli simultaneously [Landahl, 1938; 1939; 1940a]. Though the postulates are very different one may notice the similarity with Grossberg's dipole network [Grossberg, 1972a; 1972b].

Stimuli S_1 & S_2 sets into action two mechanisms working together.

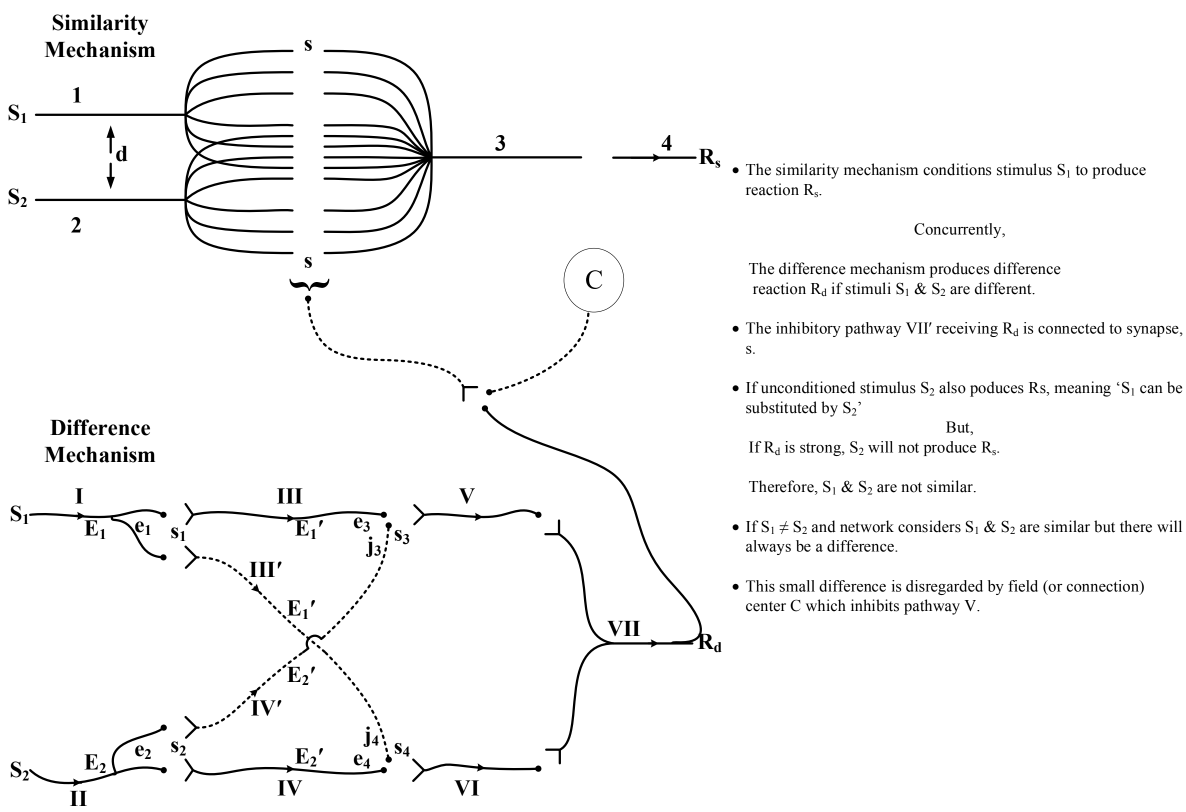


Figure 1.13. Geometrical arrangement introduced by Landahl for discriminating stimuli which evokes two mechanisms: similarity and difference. In the similarity mechanism d is parameter that corresponds to spatial distance between the sensory stimuli. The amount of s is a function of d such that the closer the stimuli are so are the number of connection between 1, 2 & 3 pathways. Center-C was not explicated and hence can be thought of as a proxy function determining the amount of acceptable small differences for similar stimuli.

Expanding on this Landahl tackled the problem of discriminating two stimuli which are temporally separated [Landahl, 1940b]. He merges ideas from the above two basic arrangements resulting in a more complicated form (see Fig.1 in [Landahl, 1940b]). The resulting arrangement has cross-coupling inhibitory pathway seen in figure 1.12c but has a single input pathway receiving simultaneously the temporally separated stimuli. By mathematical analysis Landahl concluded that spatial separation of the two stimuli is an important factor.

All the above arrangement dealt with the problem of discriminating stimuli intensities assuming they are the same sensory modality. Landahl then considered the problem of stimuli of different modalities, for instance, touch sensation from different parts of the skin. Thus for the case when two stimuli may have different intensities with same or different modalities he came up with a network system based on the hypothesis that the two stimuli set into action two mechanisms; a similarity mechanism and a difference mechanism (Fig.1.13).

The similarity reaction (R_s) from a similarity mechanism is a function of spatial distance between the stimuli. In other words, the density of connections between the branches of 1 and 2 with 3 pathway, and hence the excitation in pathway-3, is a function of the distance between 1 and 2. However, this reaction is inhibited if the difference mechanism yield a difference reaction (R_d).

Landahl further points out that no two different stimuli can be exactly identical. In other words, they may be similar but R_d occurs due to the small difference from their inequality, thus requiring a center-C which judges the amount of small difference to be neglected. Based on this, center-C inhibits the inhibitory connection from the difference to similarity mechanism.

In conclusion, the works of Rashevsky, Householder and Landahl, though modelled at a neuro-physiological level, provide great insight for possible geometrical arrangements while building neural networks modelled at the psychological level.

The neural network developed for this project is based on Stephen Grossberg's adaptive resonance theory (ART). To our knowledge the developed neural network is the first to demonstrate self-determination of features without being forced upon it.

Feature detection has been a principal task of neural networks since the beginning of the field in the late 1950's work of Rosenblatt and, independently, Widrow. Throughout the 1960's "features" were predefined by the designer of the neural network and supervised training was used to force the network to "learn" the feature. This approach is quite obviously one that, epistemologically, presumes it is legitimate to build into the network objective knowledge (since "features" are object features) a priori. Furthermore, it has long been realized by neuroscientists that supervised learning is a wholly inaccurate and inadequate model for human learning and intelligence. A baby is simply not capable of performing supervised learning.

The need to develop neural network learning systems with unsupervised learning capability was recognized at the time of the 1960's and beginning of the 1970's by Grossberg. Grossberg's Avalanche Network [Grossberg, 1969a] was the first neural network that successfully demonstrated the capacity for unsupervised learning of spatio-temporal pattern sequences. However, this network was not capable of self-determining what features should trigger its learning and had to rely upon the designer to supply it with an externally-generated triggering signal. Grossberg came to call this "ritualistic learning".

For the next 15 years, owing to a drying up of the neural network research funding after the publication of Minsky and Papert's influential book criticizing the perceptron, there were only three neural network theorists of any note carrying out research: Grossberg in the U.S., Kohonen in Finland, and Malsburg in Germany. Of these three, only Grossberg's work maintained a focus on biological and psychological plausibility for the models used. Grossberg recognized the shortcomings noted above in the Avalanche Network and was working in parallel to find solutions for them. By 1974 he was able to present a grand overview that addressed many of the key issues involved in unsupervised learning, including what turned out to be a very prescient observation that affectivity capacity in the brain was, psychologically, the most promising likely source for triggering the learning function in an Avalanche. He also came up with an empirical hypothesis that motor unit functions, properly ordered with the network, appeared to have important involvements in unsupervised pattern learning. Grossberg summarized his results in what became a landmark paper in neural network theory [Grossberg, 1974]. However, in all this work what was to constitute a "feature" in comparison and classification was still left to the network designer to define. This is again, nothing else than the injection of objective knowledge a priori into the structure of the network. Epistemologically, this is unacceptable. However, the design paradigms Grossberg developed during this period became the de factor design methodology for neural network theorists used to this day. Shortly afterward, Grossberg discovered the ART principle and attention shifted away from the still-unsolved problem of how to get a neural network to self-define what is to constitute a "feature".

There were legionary issues in the problem of what became known as feature detectors still to be solved. By 1975 Grossberg had made important advances for solving many of these issues, which he laid out in another important paper [Grossberg, 1975]. However – and this is very important – a “feature detector” was by definition a network that actually identified objects or parts of objects based upon already-predefined “features”. It did not identify what a “feature” per se was. What was to be regarded as a “feature” was still left to the ad hoc decision of the network designer.

By the mid-1980’s neural network funding had been revived, largely due to papers published by the “parallel distributed processing (PDP) school” led by Rumelhart and McClelland. Although the “new” neural network theory that emerged generated much enthusiasm, in point of fact these works did nothing else but rediscover, in incomplete forms, things already discovered by Grossberg – a fact Grossberg himself noted forcefully in another paper [Grossberg, 1987].

By 1985, Grossberg himself had turned away from exploring fundamental issues – especially the issue of how a network could be capable of self-recognizing what to use for a “feature” – to more applications-oriented applied network theory based on ART. Grossberg himself tells us that his focus had shifted in one of his books [Grossberg, p.143, 1988]. Thus, for the last quarter of a century no research effort has been devoted to returning to the “what is a feature?” problem until the work presented in this dissertation.

Below is an illustrated example that is representative of the sort of very complicated neural network systems that have been standard approaches for the many ad hoc attempts to deal with ill-posed problem issues that the inability of conventional neural networks to self-determine “what ‘features’ are” lead to. The alternative to these sorts of very complex networks is, at the time of this research, still the venerable but neurologically and psychologically unacceptable recourse to supervised learning [Grossberg, 2005].

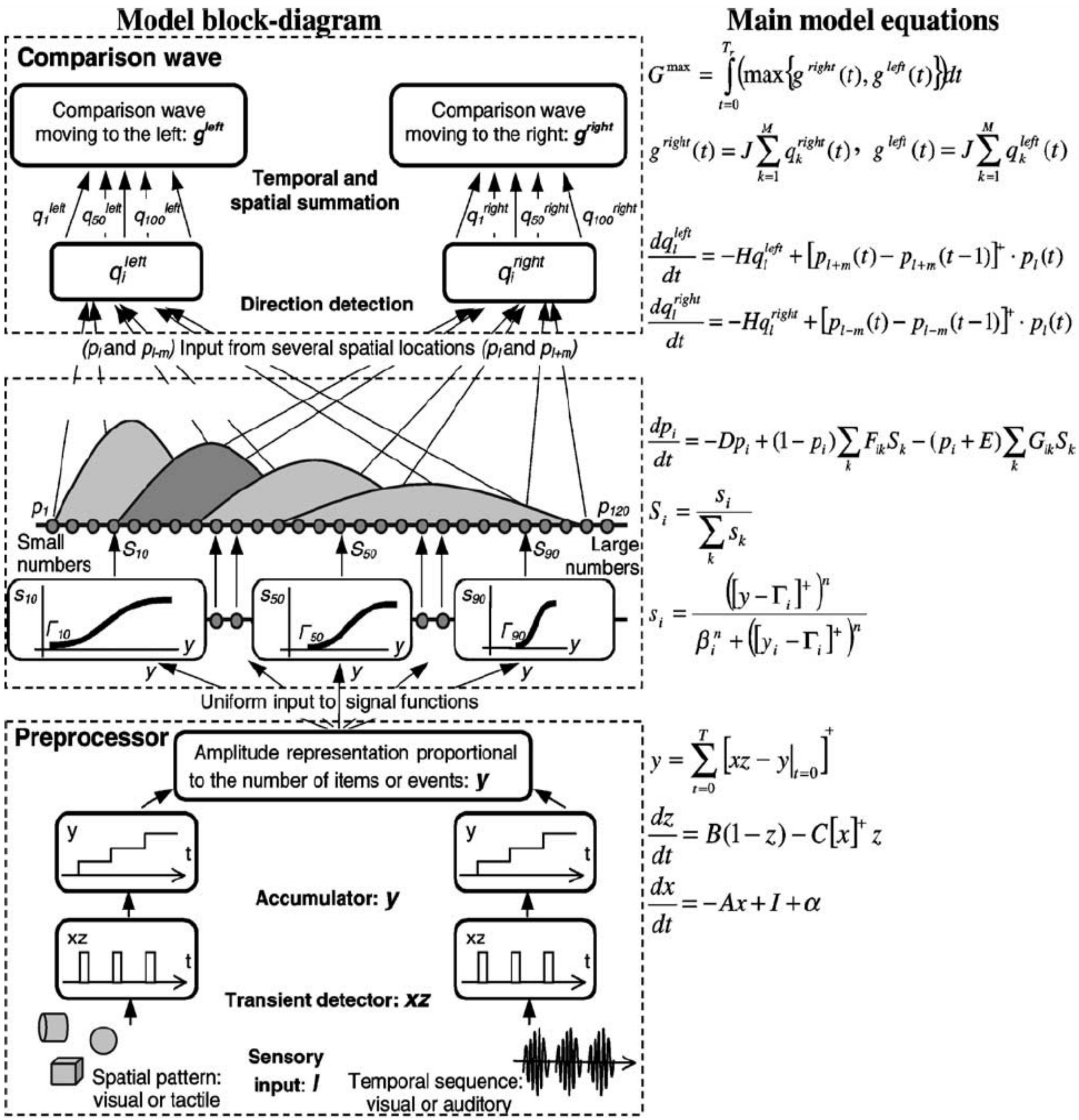


Figure 1.14. Functional diagram of the SpaN model. Preprocessor: For each sensory input (pattern or sequence), a value of the integrator y is computed at each moment in time and then uniformly fed in to the spatial number map. Spatial number map: Each activity p_i of the map receives the normalized output S_i that is derived from the same integrator input y: The signal functions s_i that give the rise to S_i have increasing thresholds and slopes at each successive map cell i. Examples for cells 10, 50, and 100 are shown on the diagram. Each 'bump' on the spatial number map schematically represents the activation pattern p_i for sensory inputs with different number of items at the moment when the whole pattern is already processed. Comparison wave: Activation of both left and right direction-sensitive cells (q_i^{left} or q_i^{right}) receives the input from two cells of the spatial number map (p_l and $p_{l\pm m}$). Both right and left magnitudes (g^{left} or g^{right}) at each moment are computed as summed activations of corresponding direction-sensitive cells (q_i^{left} or q_i^{right}). [Grossberg, 2003]

Grossberg and Repin in their 2003 paper [Grossberg, 2003] introduces a neural network for the problem of numerical representation (in brain) and their comparison. They called their network Spatial Number Network (SpaN). It is functionally made up of: preprocessor, spatial number map and comparison wave (Fig.1.14).

At preprocessor stage pattern inputs presented spatially or temporally are given a numerical input (dc) value for each respective pattern. An input induces an activity level. The activities are accumulated proportional to the number of presented patterns.

The accumulator value passes on to the spatial number map which is made up of k-nodes. The accumulator is transformed to normalized input for the spatial number map. This is done such that, map nodes on the left hand side gets activated faster than those on right for a given accumulator value.

The spatial number map output is then transformed to direction-sensitive (left & right) activities. They are respectively added to form comparison waves, moving to left and moving to right. For two pattern inputs, if amplitude of left-wave is greater, then the second input is smaller. On the other hand, if amplitude of right-wave is greater, then the second input is larger.

Though the SpaN demonstrates capability of comparing patterns, it is different from our neural network for comparation. First, patterns are assigned numerical values in SpaN. In other words, features are determined beforehand. Second, the on-center, off-surround property of nodes in the spatial number map is incorporated by using normal-distribution equations for excitatory (F) and inhibitory (G) Gaussian kernels.

This ends the background discussion. The next chapter demonstrates the subtleties in the task of comparison, in other words, things we take for granted in our everyday performance of comparison.