Preliminary Discussion of the Martian 2 Program

I. Introductory Background and Aims of Martian 2

A Martian is an artificial intelligent agent that is designed based on the principles of mental physics. The Martian Research Program began in 2008 with the Phase I Martian Infant Model, which I will hereafter refer to as Martian 1. The goal of Martian 1 was to explore neural network system architectures capable of realizing an agent that produced observable behaviors congruent with those of human infants in the first month of the sensorimotor stage of psychological development [Piaget (1970), chap. 1]. The Piagetian model was chosen because Piaget's theory, alone among all those posited to date by empirical psychology, is the only one extant that is (1) systematic across the spectrum of human behavior; (2) is epistemology-centered; (3) is congruent with the theorems of mental physics [Wells (2009)]; and (4) is so far unrefuted by objectively valid conclusions drawn from properly conducted experiments. The specific objectives of Martian 1 focused on neural network system research, but the overarching aim was to gain concrete experience for understanding how the context of neural network theory could made to fit the requirements of mental physics. Wells and MacPherson (2009) documents the outcomes of this research project. The program was primarily funded by a National Science Foundation (NFS) Research Experience for Undergraduates (REU) site grant, in partnership with financial support from the Air Force Office of Scientific Research, and partially supported by the Idaho Engineering Experiment Station. In the aftermath of the U.S. recession that began in 2008, NSF cut the funding for half of its REU sites and funding for the program was lost. Since that time, both NSF and the National Institute of Mental Health have declined to fund the program, and so the Martian 1 program was forced to be terminated in March of 2010.

The experience gained from Martian 1 was important because it helped bring about a concrete understanding the approach necessary for solving the *basic* research problem of expressing a mathematical mental physics doctrine in such a way as to lead to a correct mathematical theory of mind-body natural science. Specifically, Critical analysis of what was learned from Martian 1 contributed to making a breakthrough discovery: the long-sought secret of the neural code, which is now understood to be not merely a neural but, rather, a *somatic* code. This analysis also brought forth a new science, Critical semantics, which is the name I have given the mathematical part of the empirical science of Critical psychophysics [Wells (2011a, b, c)]. The effort to develop this new science also demonstrated the need to formally develop a Critical doctrine of method for developing *any* Critical applied metaphysic, which is necessary for a solid grounding in *every* natural science [Wells (2011d, e)]. Additional possibilities opened up by this are impressive: (1) the transformation of every existing social science into a true *social-natural* science as capable within its scope of applications as physics is within its scope; (2) the founding of a new social-natural science of *leadership* [Wells (2010)]; and (3) laying the groundwork for a true social-natural *science* of education.

The key findings to come out of this recent work in regard to mind-body science are: (1) the theory of the somatic code is a topological theory; (2) this theory is at its roots the mental physics of social-natural semantics expressed in the mathematics of a general field theory[1]; and (3) the expression of this theory is made by a unification of embedding field theory and Critical topology theory. The theory of the somatic code realizes a goal expressed by Weaver in 1949 and extends Weaver's original concept well beyond just communication theory [Wells (2011f)].

[1] By "general field theory" I do not mean that part of the mathematics of abstract algebra called field theory. Rather, I mean generalized field theory as presently exemplified by the physics of electromagnetism and by quantum field theory. Critical field theory is a relativistic quantum field theory, of which physics merely provides some specific examples. Biological receptive fields are underdeveloped examples.

Martian 2 is a continuation of Martian 1 but taken in a new and better direction. The specific aim of the program is the same as Martian 1. However, its method has changed significantly and the program aims at developing new *general* practices useful for (1) advancement of the arena of scholarship currently called computational neuroscience, (2) transformation of empirical psychology into a social-natural science, and (3) engineering research in what is called the field of computational intelligence. Mental physics teaches, as a *theorem*, that human beings learn *from* the specific *to* the general, and only after more general and abstract concepts are developed via this route is it possible for a human being to then apply those general concepts in deduction and application to other topics. Martian 2 research, as intelligent agent research, is a vehicle for obtaining this sort of specific scientific experience. An intelligent agent system – which is an engineered system – is a more prudential approach to the general problem than a direct attempt to apply Critical psychophysics to current neuroscience would be. This is because an intelligent agent can be designed to capture human-like behavior and development characteristics and thereby uncover functional invariants likely to be found operating as well in the human central nervous system (CNS). A direct application to the CNS at this time is one that would have to overcome an enormous hindrance, namely, the unimaginable complexity of the CNS. This complexity underlies our attending lack of sufficient knowledge of CNS anatomical and physiological functions to positively establish *with objective validity* what functions different CNS anatomical structures actually implement in H. sapiens. Hypotheses are not facts.

I have contended for many years that biology and psychology can learn much from the science of engineering and that engineering can learn much from biology and psychology. This point of view has been a precept of the Martian program since its inception. I also contend there is an unappreciated symbiotic unity inherent in what is known by the now-popular acronym STEM (Science, Technology, Engineering, and Math). In its popular form, STEM has not been with us for very long and, in my opinion, that idea is just a vague notion, an ideal and a hope held by its proponents. It lacks *practical* visions and plans for the future of public education. Numerous conversations and interactions I have had with colleagues, my participation in a recent NSF panel, and observation of the current STEM Education advertising campaign, lead me to conclude that most educators do not distinguish between "technology" and "engineering" and actually view STEM only as what would have to be tagged STM.

Many of the colleges of education faculty members I know appear to not understand what "engineering" is. They think E = T and most of them relate to technology by no more than a relatively trivial view of computers-in-the-classroom, Internet-based education tools, and similar gadgetry. These are useful gadgets, of course, but still merely gadgets. Gadgetry in and of itself solves no social problems and is easier to misuse than to beneficially employ. Harsh as this criticism may be, I have been a STEM researcher for nearly four decades and I think my criticism of the current state of STEM enthusiasm is amply justified. It will take a true social-natural *science* to bring unity to STEM education, and without that unity STEM will become merely another flash-in-the-pan trendy fashion that will hang on for a few years and then join the host of other nobly-inspired dead programs I have seen flower then fade over the years. In my opinion, the Martian 2 Program will likely contribute indirectly to bringing that unifying social-natural science into actuality through collateral discoveries I anticipate it *will* bring to light.

This broader vista is a longer term aim. Experience strongly suggests that attempting to leap straight to this science without producing scientific stepping stones to reach it is far more likely to fail than to succeed. For that reason, the specific aims of Martian 2 are much nearer term and quite specifically engineering focused. At the same time, though, the Martian 2 program can not be made so tightly focused as to lose its necessitated connections with neural science and psychology. Maintenance of those linkages sets the direction for the requirements of Martian 2 anatomy and for the mathematical methodology to be employed in actualizing a Martian agent.

II. Fundamental Architectural Considerations

The engineering design of any Martian system architecture is constrained by definition. To meet the longer-term goals of the research, a Martian must be an agent system whose functional organization is a *soma*-like small-scale mimic of H. sapiens and whose functions instantiate the processes and capabilities of the phenomenon of mind that mental physics explains. Although it is a somewhat poor simile to use, the mental organization of H. sapiens can be likened to the operating system structure of a machine with the agent-system standing in the role of the body phenomenon. The Martian agent and every subsystem within it must be designed under the constraint that it be possible to link its structure, through Critical semantics and the somatic code, to the mental organization and structure of *nous*. Martian design is always an essay in Critical psychophysics at the smaller scale of a modest abstract image of H. sapiens. This overarching architectural plan is illustrated in figure 1, which presents the general organization of the Organized Being (H. sapiens). This structure is the direct outcome of the mental physics of the phenomenon of mind [Wells (2009)].

I present this figure without apology and with full cognizance that very few people will really grasp its meaning at the time of this writing. I will also be using the technical language of mental physics throughout this paper with this same cognizance. This is a technical paper and a working document of the Martian 2 program, and I cannot teach mental physics in these pages. Where I can, I will do what I can to help the novice reader follow the gist of the theory. This paper is, after all, primarily aimed at people who will work or who are working on the Martian problem, and they are not trained mental physicists (yet). The theory is simply too new (5 years old this upcoming September) for it to have become widespread knowledge. As aids to the novice: (1) the technical language is provided in a convenient glossary [Wells (2011g)]; (2) the textbook is *The Principles of Mental Physics* [Wells (2009)]; and (3) the foundational research is covered in Wells (2006). All of these are publicly available via my Internet homepage at no cost.

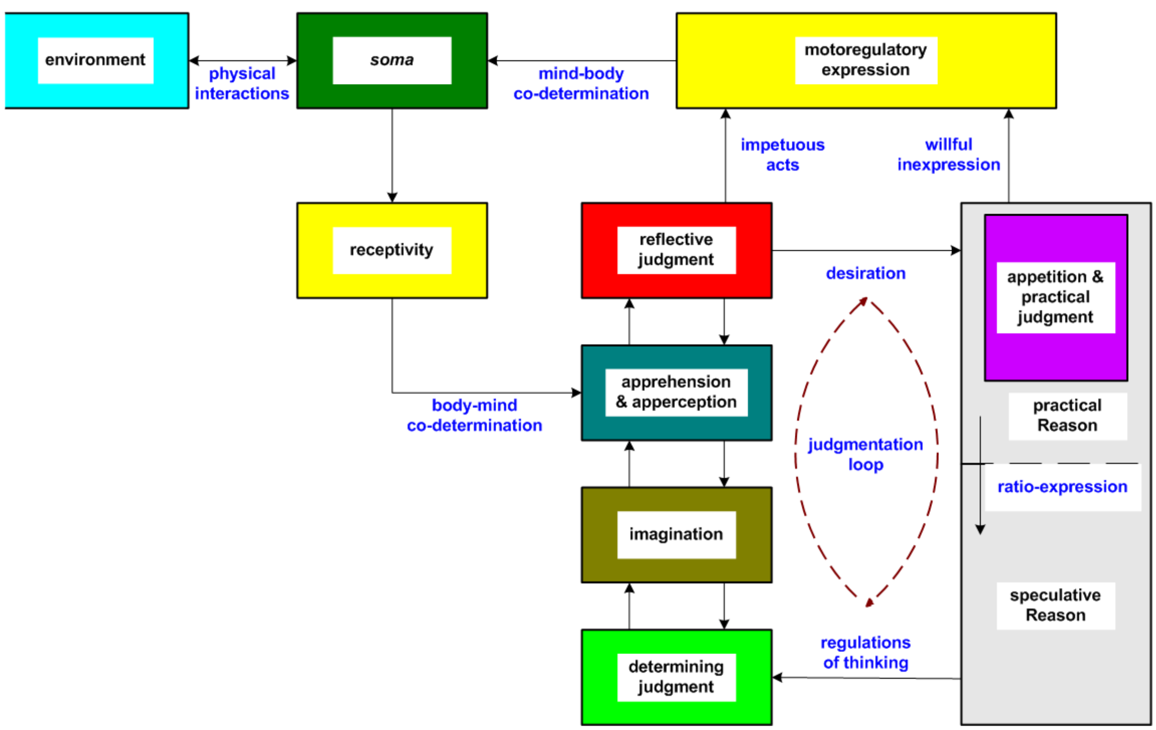

Figure 1: Organization structure of an Organized Being.

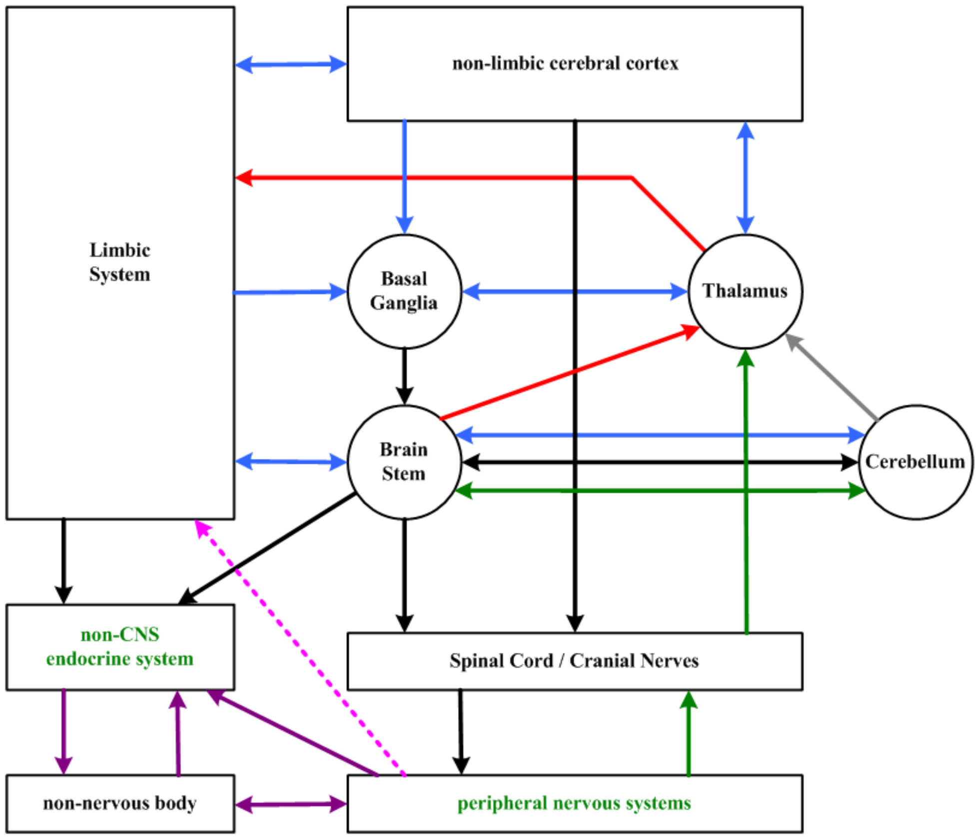

Figure 2: Extended Brooks' model of somatic architecture.

A Martian system is a physical model of *soma* for a simplified Organized Being. This is to say it is an image projection of the body phenomenon of H. sapiens. If a Martian is to be in any context regardable as a proxy for H. sapiens, then its architecture is constrained *a priori* to be such that it comply with the requirement for homeomorphism between somatic semantics and noetic semantics as discussed in Wells (2011c). It was the lack of a precise understanding of this mathematical constraint when Martian 1 began, and the achievement of this understanding now, that underlies the change in technical methodology in going from Martian 1 to Martian 2. This is to say that Martian 2 has a more precise *context* than Martian 1.

The immediate objective for Martian 2 has not changed from Martian 1. Broadly stated, this objective is to understand the functional requirements of organized being that ground behavioral and intellective characteristics exhibited by human infants in the sensorimotor phase of child development. The program's evaluative standard is not whether the Martian as artificial agent can do this or that engineered task; it is to judge the extents to which it does and does not exhibit behaviors congruent with known findings of human behaviors and capacities at well-defined points of human psychological and physical development. The technical *poiēsis* is also the same, namely, the method of minimal anatomies first stated in 1972 [Grossberg (1972)].

Because minimal anatomies is the pragmatic methodology, it seems by a commonsense view of the matter at hand that a somatic architecture for a Martian ought to be one closely allied with the scientific knowledge in our possession in regard to human anatomical organization. Further, there is a known organizational schema, first proposed for sensorimotor system organization by Brooks in 1983 [Brooks (1986), chap. 2], that can be further extended for employment as an architectural schema of mammalian somatic organization. Figure 2 illustrates Brooks' model.

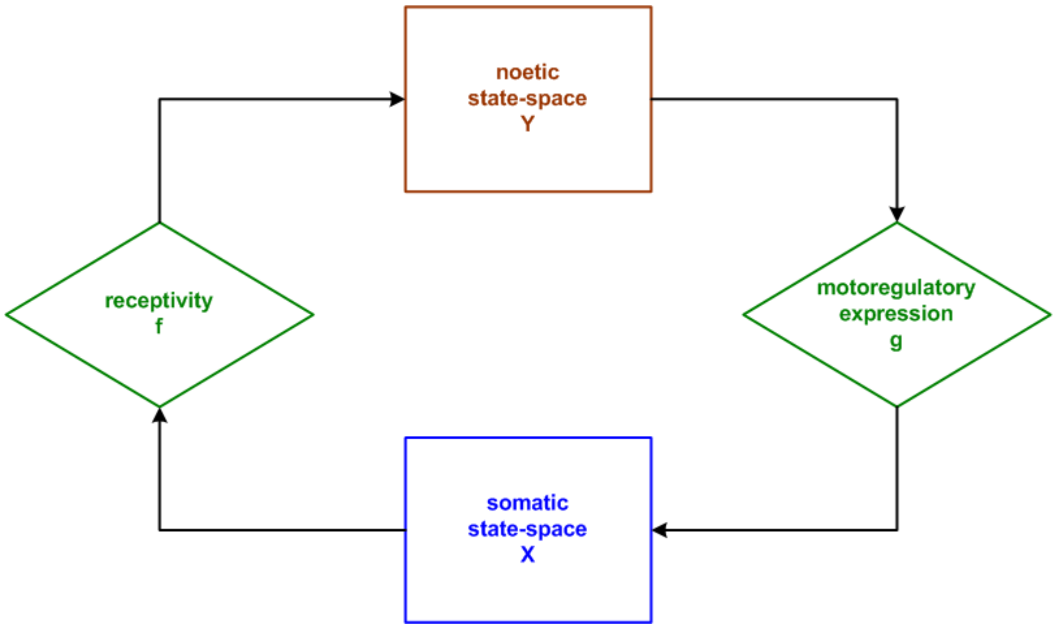

Figure 3: Noetic and somatic semantic state spaces with homeomorphism mapping by *psyche*.

But when we compare Brooks' model with the mental organization in figure 1, we find a major twofold problem inherent in this commonsense idea. First, there is not a one-to-one mapping from the structures in figure 1 to those of figure 2 or *vice versa*. This is not a surprise because it has long been known to neural science that the neural correlates to psychological phenomena, as observed through brain imaging techniques, are widely distributed throughout the CNS. This means that the corresponding structures for homeomorphism are non-obvious between the somatic-state and noetic-state semantic spaces (figure 3) [Wells (2011c)]. The anatomy of Martian 1 took its architectural theme directly from Brooks' model, but the experience gained from that project showed that getting from Brooks' model to a homeomorphic mapping into figure 1 is impractical. It also had not been known initially that a homeomorphism is *necessary*.

The second part of this twofold problem is likewise not surprising. At the present time, neural science simply cannot supply enough anatomical detail to systematically reduce Brooks' model to more specific CNS representations in order to more precisely depict anatomical functionalities. The issue here is less severe in degree regarding the non-limbic cerebral cortex, limbic system, cerebellum, thalamus, basal ganglia and the ventral horn of the spinal cord. There are more detailed anatomical data available that lets these large-scale blocks be reduced to more specific functional blocks (although there are limitations soon encountered even here). The dorsal horn of the spinal cord presents a greater challenge, but the major shortcoming lies in lack of functional organization knowledge about the brainstem. Its major anatomical divisions (i.e., medulla oblongata, pons, midbrain, and details of the reticular formation) are readily identified by neural science, but finer level details of intra-brainstem connectivity expressed *systematically* fall short of the level of detail available for the other blocks in Brooks' model. This is not to say there is a lack of data concerning particular aspects of, e.g., the superior or inferior colliculi [Hall and Moschovakis (2004), Winer and Schreiner (2005)]. It *is* to say this data has not been systematized and is scattered piecemeal across the neural science literary *corpus*. This is a serious matter because it *is* clear that brainstem functionalities have functional correspondences with the process of reflective judgment, the synthesis of empirical apperception, and the synthesis of apprehension in figure 1. Another, no less serious, issue is that the theory of the somatic code clearly tells us that the *endocrine* system *cannot* be omitted from the model.

I am forced to conclude, then, that it was a mistake in Martian 1 to take a somatic basis for approaching the agent architecture problem. It will likely not surprise him, but analysis of Martian 1 leads me to conclude that Grossberg's general approach in 1968 is the most practical one to take for efficacy of architecture *identification*. His method [Grossberg (1968)] begins with psychological postulates and proceeds from them to mathematical requirements and then to what he later came to call "mock anatomies." The only difference between this method and the method to be used in Martian 2 is that Martian 2 will take its postulate bases from the theory of *nous* provided by mental physics and use empirical findings of psychology as *tests* of the empirical postulates of somatic architectures and anatomies. The norm for evaluation here is twofold: (1) Martian 2 behaviors must be congruent with empirical facts; and (2) the Martian 2 somatic architecture must be such that homeomorphic functions [f, g] in figure 3 can be obtained. Our present state of understanding of neurology is simply too sparse to *rigorously* support the Martian 2 architecture identification problem. There is simply too much modeling error.

Nonetheless, this does not leave Brooks' model without a role. Semantic field representations still have to be associated with somatic field representations [Wells (2011c)] at some level of set membership representation in order to provide mathematical *principal quantities* [Wells (2011e)] necessary for objectively valid empirical concepts of *soma*. Furthermore, as mock anatomies are systematically deduced for Martian 2, we can expect these to yield only partial embedding field spaces and not the universe of topological spaces in the beginning. *Coordination* of the divers partial embedding field spaces so posited seems very likely to benefit from whatever connectivity information we can glean from known findings of neurology. A *systematic* architectonic can still be *guided* by Brooks' model even if this model cannot present a distinct enough overall picture of somatic architecture. Just as the theory cannot be divorced from empirical psychology, so also it cannot be divorced from empirical neurology and neural science. All that has happened between Martian 1 and Martian 2 is that neurology and psychology have exchanged places in deductive priority in supporting the research program. Brooks' model remains the principal empirical support for the *empirical* anticipation that large-scale homology will emerge as an architectural schema in regard to divers sensorimotor modalities. Furthermore, it provides an initial basis for establishing the all-important requirement for *context* in positing *theoretical* semantic fields.

III. The Schema of Noetic Functions

The field mechanics of somatic-state physics are represented in embedding field theory by arc functions called *associational strengths* that establish connections between places in somatic state space in a graph-theoretic mathematical form [Grossberg (1969, 1971), Wells (2011c)]. It is by means of these connections that semantic representations are synthesized in the learning process. This is universally recognized in present day neural network theory, although theorists tend to be only semi-conscious of the semantic context of neural network graphs (with the notable exception of Grossberg and others familiar with embedding field theory). It is, however, reflected by the terminology of the science in such phrases as "exemplar vectors," "classifications" or "chunking."

Objectively valid understanding of the semantic context portrayed in neural network graphs requires us to take a closer look at the word "associational" in "associational strength." What does this adjective mean? The primary technical definition of "association" in mental physics is *the function of aesthetic Quantity in producing a relationship of commonality for two or more representations in conscious presentation* [Wells (2011g)]. An association makes a combination of two or more things. In the mental physics of *nous* one way combinations are synthesized is by a process called *judgment*. To understand the synthesis of semantic context presented in embedding field graphs requires us to understand the mental physics of combination processes, which is to say it requires us to understand what minimal anatomies are necessary to synthesize combinations in acts of knowledge representation.

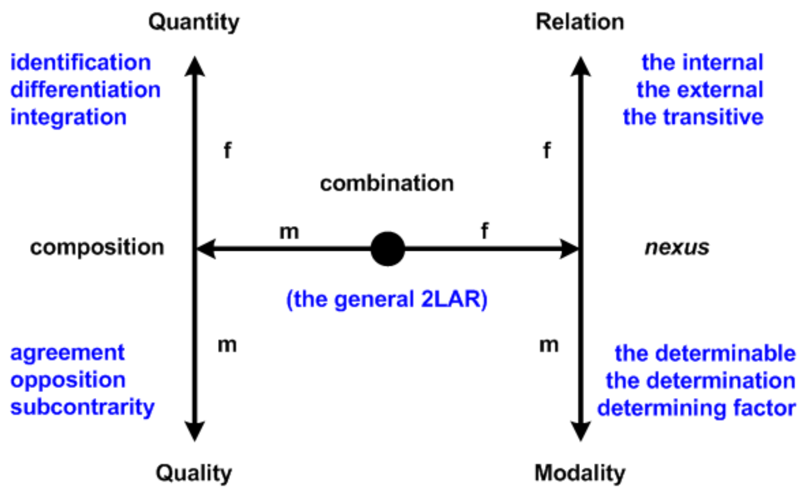

Figure 4: Second level analytic representation (2LAR) of the synthesis of combination.

Mathematical combination theory in knowledge representation is outlined in Wells (2011b) and explained in more detail in Wells (2009). All specific acts of synthesizing a combination are species under one genus of representation illustrated in figure 4 [Wells (2009), chap. 2]. The twelve terms named in figure 4 are general functionals[2] and are called the **general *momenta* of combination**. Every special act of combination is special by virtue of its 2LAR *momenta* being specialized for the context of what that act of combination depicts. Every act of combination is a function comprised of four *momenta* drawn from the specific set of functions pertinent to the four headings depicted in figure 4. For example, a specific act of combination might be synthesized as

combination = {integration, opposition, transitive Relation, the determinable}.

There are, therefore, precisely $81 = 3^4$ *general classes* of functional combination.

The total number of *specific* combination functions is, of course, much larger than this because specific acts of combination in *nous* use *momenta* specific to the type of combination relation[3] being enacted. For example, the process of reflective judgment (see figure 1) contains 13,122 specific kinds of possible combination relations[4]. Although the total number of possible specific combination relations is somewhat staggering, there is a systematic unity in the Nature of all acts of combination (namely, figure 4) and it is the ability to *exploit* this systematic Nature that makes mathematical embedding field theory a practicable science. To use an analogy, the number of computer programs that *could* be written in any particular computer language is unlimited, but the systematic *structure* of that language makes computer *programming* possible. It is not without reason that a "syntax error" in a line of computer code is called a "syntax" error. Figure 4 depicts the systematic structure of combination in acts of representation. Martian 2 research involves the exploitation of this systematic structure.

To further illustrate this point, consider for specificity the process of the synthesis of apprehension and apperception (see figure 1) depicted in more detail by figure 5 [Wells (2009), chap. 3 §1]. Excluding the acts of reflective judgment, imagination, and determining judgment, there are five specific processes of combination found in sensibility.

---

[2] A functional is a function that has a domain that is a set of functions and a range belonging to another set of functions [Nelson (2003)].

[3] An *n*-ary relation is an association between or property of *n* mathematical objects [Nelson (2003)].

[4] This is a theorem of mental physics. One should note that "relation" ≠ "Relation" in this terminology.

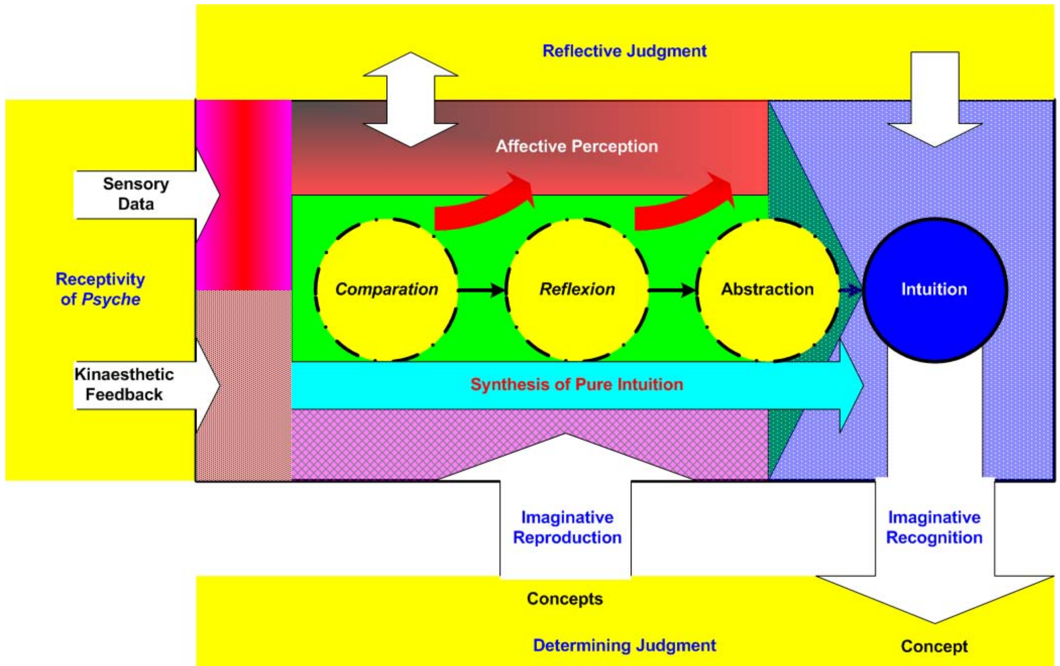

Figure 5: The synthesis of apprehension and apperception in sensibility.

The first three of these are jointly called the *Verstandes-Actus* or "acts of understanding" in sensibility. The *Verstandes-Actus* are the syntheses of *Comparation*, *Reflexion* and abstraction. *Comparation* is the synthesis of mathematical equivalence relations [Wells (2009), chap. 11 §7]. *Reflexion* is the synthesis of congruence relations [*ibid*.]. Abstraction is the synthesis of selective negations, a form of negative association for which mathematics does not have a specific name. The *set difference* operation in formal set theory is an example of a relation of abstraction. The difference between two sets, A and B, is denoted in mathematics as $A \backslash B$ and is formally defined

$$A \backslash B = \{x: x \in A \ \& \ x \notin B\}.$$

The synthesis of abstraction takes an associated set A from the synthesis of Reflexion, partitions it into two sets, A and B, and then performs the set difference operation indicated above. In the language of embedding field theory, this is called ***quenching*** an activity [Grossberg (1973)].

The remaining two synthetic processes are implicit in the synthesis of pure intuition depicted in figure 5. They are called the *synthesis of the pure intuition of subjective space* and the *synthesis of the pure intuition of subjective time* [Wells (2009), chap. 3]. The synthesis of the pure intuition of space is a synthesis of topology structuring. The synthesis of the pure intuition of time is a synthesis of mathematical partial order structuring [Nelson (2003)]. Thus, all five processes make up the overall process of the synthesis of apprehension in *nous*.

The synthesis of apperception, on the other hand, is a synthesis of an interaction between the acts of sensible representation and the process of reflective judgment. To *judge* is to subsume particular representations under a general representation. The synthesis in sensibility performs no judgments. Rather, it synthesizes the matters upon which processes of judgment act. In a manner of speaking, the synthesis in sensibility synthesizes many *potential* representations and presents them to reflective judgment. The latter partitions them into three classes: (1) affective perceptions; (2) intuitions, which are objective perceptions; and (3) obscure representations, which are representations that are not transformed into perceptions. Perception in general is representation with empirical consciousness. Empirical consciousness is "the representation that another representation is in me and is to be attended to" [Wells (2011c)] and is *presented* as a representing mark affixed by the process of reflective judgment working in conjunction with the depictions produced in the synthesis of sensibility. In the language of embedding field theory, functions that carry out this presentation are called attentional and orienting functions [Grossberg (1975)].

One way to look at neural networks is to look at them as functional mechanisms synthesizing embedding fields under the influence of external aliments ("inputs") that they act to either assimilate or negate. The umbrella research question for Martian 2 is: "What *kind* of embedding field?" Put another way, a working neural network anatomy *will* produce an embedding field of some sort. How does one know it is *that* embedding field, and not some other, that was the real object of the anatomy one has built? If one already knows the *context* of his design work (e.g. "I want to build a classifier"), this is not much of an issue because there is more than forty years of well-honed mathematical theory already in the design toolbox, much of it due to the work of Grossberg and his colleagues and students since the late 1960s. But how does one decide that it is a "classifier" (for example) that one wants to build? How does one know that the "classifications" it produces are ***actually*** pertinent to the phenomenon of being a human being?

That is a much trickier question. It is characteristic of all system theory and all system design that the problem of figuring out *what* you want to build is much more difficult than the problem of figuring out *how* to build it (or, at least, it is for a well-trained engineer). Martian 2 has a very specific Object of research, namely the phenomenon of what is required for an object *Existenz* to be a ***human*** *Existenz*.[5]

Here is where the general schema of combination (figure 4) and the mental physics of the phenomenon of mind enter the picture. I don't doubt for a second that a great many handy little design tricks ("techniques") will be discovered or *re*-discovered simply in exercising the method of minimal anatomies in developing the Martian 2 agent. I don't doubt that interesting new applications for already-known neural network anatomies will be discovered (or re-discovered) in the process. I don't doubt that *new* neural network anatomies will be discovered. All these things also happened during the Martian 1 phase. The ***most significant contributions*** to new knowledge will come from understanding the Critical and the mathematical bonding between embedding field constructs and the various semantic schemata of Critical representation that will unify the theory of the phenomenon of *body* with the theory of the phenomenon of *mind*. *This* field of science presently lies fallow and is awaiting planting and harvesting.

IV. Homeomorphisms, Anatomical Architectonic and Observability

The example of the previous section, the synthesis in sensibility, serves to illuminate why I said earlier that Brooks' model was unsuitable as a basis for architecting the Martian 2 agent. I have described the synthesis functions in this process. Now let us examine how what we know about neurology and physiology are related to them at the level of Brooks' model.

I mentioned earlier that the findings from, e.g., brain imaging experiments has demonstrated that neurological correlates of psychological phenomena are widely distributed throughout the CNS. Let us look at how the anatomical blocks in Brooks' model stand in relationship to the synthesis of apprehension and apperception.

[5] *Dasein* is "existence" in the connotation "something exists." *Existenz* is "existence" in the connotation of *how* that something exists. To posit a *Dasein* is merely to posit that there is some object; to posit *Existenz* predicates is to *represent* that object. To say "Zeus is" is to say almost nothing. To say, "Zeus is my neighbor's dog" is to tag "Zeus" with a specific meaning – and *that* is an act of Critical semantics.

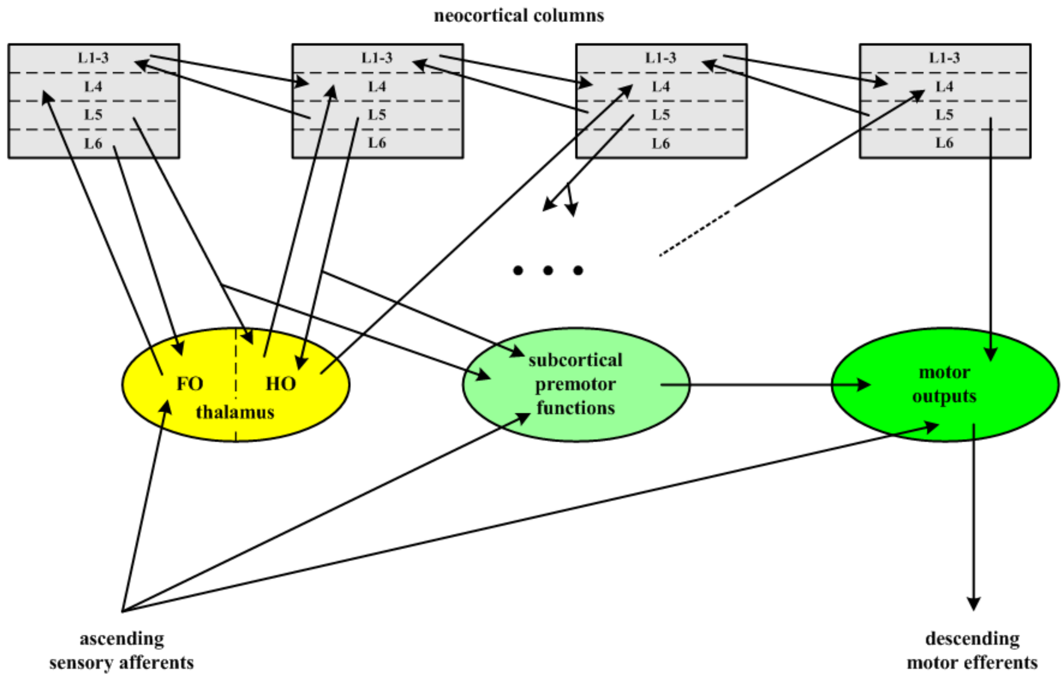

Figure 6: Simplified illustration of thalamocortical connections.

In the first place, we know that the boxes labeled spinal cord/cranial nerves, brainstem, thalamus, limbic system and the motor cortices in the neocortex are in some kind of relationship with the synthesis of sensibility in *nous*. One fact that tells us this is that these are involved with the synthesis of the pure intuition of space as well as with the synthesis of perceptions. We know this because the mental aliments for the synthesis of the pure intuition of space are the kinaesthetic feedback data of receptivity depicted in figure 5 [Wells (2009), chap. 3]. The transcendental place of origin for kinaesthetic feedback is in *soma* and is exhibited in *soma* by sensorimotor corporal structures that include, but are not limited to, the biological mechanisms of eye saccades, the skeletal muscle feedback from intrafusal muscle fibers and the mechanisms of the vestibular system.

Now in addition, sensory afferents and thalamocortical connections also involve cortical and subcortical *motor* functions. A simplified illustration of this is shown in figure 6 [Sherman and Guillery (2006)]. This means these are *also* involved *at a minimum* with reflective judgment and with appetitive power in practical Reason. This is because motoregulatory expression expresses noetic information that is produced in reflective judgment and in appetition [Wells (2009), chap. 10]. In the mappings between figures 1 and 2, we therefore have multiple projections from blocks in figure 2 into blocks in figure 1 and *vice versa*.

This means that these boxes in Brooks' model make one-to-many projections into the structures in figure 1. But a homeomorphic function (in particular, the functions f and g in figure 3) *can not make* one-to-many projections from somatic state space to noetic state space or *vice versa* and, in fact, the mappings must both be one-to-one. We know that the homeomorphic pair [f, g] in figure 3 is necessary for the possibility of meaning implications in *nous* [Wells (2011c)] and we know that human beings make meaning implications [Piaget and Garcia (1987)] (as if you didn't already know this from your personal self-experience). Put these facts together and a conclusion follows readily: If we grant that Brooks' model is correct[6], *there must exist another*

[6] If Brooks' model was "too wrong" we wouldn't want to use it in the first place.

functionally equivalent form of somatic architecture depiction where the mappings from somatic state space to noetic state space and vice versa *are one-to-one* and, to the extent Brooks' model is correct, this equivalent somatic state space must be obtainable as a transformation of Brooks' model. If we did begin with Brooks' model, we'd just have to transform it to this other form eventually in order to establish the required homeomorphism. The method of minimal anatomies has the virtue that one finds the minimal anatomies first and later can search for their assignments to more specific anatomical substrates or *place* them in a Brooks-like model.

If we begin with the mental structures of *nous* and *psyche* instead, then we know the homeomorphic function g (because designing the minimal anatomies is at the same time the definition of this mapping). Unless the embedding field neural network subsystem has a mathematical property system theorists call an *unobservable form* [Chen (1970), chap. 5] we can also take the somatic representation straight back into the noetic one – thus ensuring we have the function f as well. This provides all the compelling reason one could ask for to justify the change in architectonic method in going from Martian 1 to Martian 2.

This does, of course, raise the issue of observability in neural network systems. Observability is a property that has not received much study by neural network theorists and, indeed, a formal closed-form treatment of system observability has only been developed for linear-and-time-invariant (LTI) systems [Chen (1970), chap. 3]. Simply put, a system is observable if and only if observations of its inputs and outputs in objective time provide sufficient information to allow the states of the system to be determined. For LTI systems this is a more or less straightforward exercise in linear algebra [Kailath (1980), chap. 2, pp. 80-84]. For non-LTI systems[7] the mathematics problem is much more difficult to solve and without the use of set membership theory (SMT) might be entirely unsolvable (although we actually possesses no *theorem* to this effect). We know of at least some systems that, treated using classical "point solution" mathematics, are unobservable but which *are* observable when treated using SMT.

The reason for this is that input, output and state variables serving as principal quantities in a SMT description are solution *sets* rather than point solutions [Wells (2011e)]. If two different "point solution" state sequences are both solutions for the same sequences of inputs and outputs, the system is unobservable in point-solution mathematics. However, two such state sequences treated by set membership techniques *belong to the same solution set*, because they are *empirically indistinguishable*, and the system is therefore *set membership observable*. This is how/why SMT techniques are able to solve problems in system identification and parameter estimation that conventional mathematical treatments find unsolvable [Walter and Piet-Lahanier (1990), Milanese and Vicino (1991), Combettes (1993), McCarthy and Wells (1997)].

SMT techniques have been demonstrated to work with linear-and-time-varying (LTV) system problems [Rao and Huang (1993)] and with non-linear-time-invariant (NLTI) system identification problems [McCarthy and Wells (1997), Brennan (1998)]. NLTI systems that are expressible by a finite Volterra series expansion [Schetzen (1980)] can be dealt with in a straightforward fashion by SMT identification and estimation algorithms [Walter and Piet-Lahanier (1990), McCarthy and Wells (1997)]. NLTI systems with saturating nonlinearities or "dead zones" typically are not expressible in Volterra series form, but I have supervised student term projects in the Wells Laboratory where SMT has successfully solved non-trivial system identification or parameter estimation problems of exactly this type [unpublished results].

Neural networks used in embedding field theory are almost always nonlinear systems. While

[7] It should be noted that when most people refer to "nonlinear" systems, they often mean "non-LTI" systems. A system can be nonlinear and time-invariant, nonlinear and time-varying, linear and time-varying or linear and time-invariant (LTI). "Nonlinear" is a term often applied to mean any of the first three types. This is, however, improper terminology and the correct term to use is "non-LTI."

they are undergoing adaptation they are also time-varying systems. However, the mathematical structure of these systems is usually partitioned into two pieces: (1) a NLTI network system such as a shunting node network [Grossberg (1973)]; and (2) a weight adaptation subsystem, the dynamics for which are sometime LTI and sometimes NLTI. It is the combination of these that makes the overall system NLTV, but because the problem can be partitioned as I have just described, the results cited above are pertinent to them.

It is worth mentioning that mathematical techniques developed in Martian 2 for analyzing the agent are likely to have some collateral benefit when system theoretic methods come to be employed in studying biological neural networks. It is already known that biological neural networks are NLTV; the Hodgkin-Huxley model is proof enough of that. It should, however, be noted that the Hodgkin-Huxley model is one of those systems belonging to the partitioned class of models and, indeed, many popular neural network models, including the shunting-node model, belong to the same class of mathematical functionals as the Hodgkin-Huxley model [Carpenter (1980), Grossberg (1980)]. Furthermore, there is compelling reason to think that biological neural networks fall into the class of point-solution-unobservable systems. This reason comes from the phenomenon of cell death. It has long been known that brain cells die at rate that is relatively rapid in comparison to the average lifespan of a human being. Abeles uses a cell death rate estimate of on the order of 0.5% of cortical cells per year [Abeles (1991), chap. 6].

If biological neural networks did not have significant *redundancy* in their structure, none of us would live past childhood. The fact that most of us do (in developed countries) is singularly good empirical evidence of redundancy in biological neural network systems. But *all* redundant systems are unobservable in point-solution mathematical models. By definition, redundancy *means* that input-output characteristics are unaltered by failure of a part of the system (up to a limit where redundancy begins to become inadequate), and to say this is to say the system is unobservable. This sort of system structure is one for which SMT provides well-posed solutions for observability. It also provides a real alternative in place of the statistical/thermodynamical mathematical foundation that von Neumann conjectured would be needed to properly develop a logic of automata [Neumann (1948)] and does not require the introduction of probability distributions, which are *never* anything else than mathematical *secondary* quantities, as *principal* quantities of the theory. Probabilities can never be used *with objective validity* as *principal* quantities in any theory of nature [Wells (2011e)].

V. Some Conjectural and Transitional Closing Remarks

In broad strokes, this paper outlines the approach principles and general tactical plan for the Martian 2 Program. It will begin with where Martian 1 ended – the first month of the sensorimotor stage of the human infant – but will re-cast Martian 1 into the form of the new method that I have described here. I do not anticipate this stage will take much longer to complete. From there the natural progression is to stages 2, 3, etc. of sensorimotor intelligence. There is a good empirical data set for these stages available for testing and evaluation purposes, and mental physics tells us Piaget was correct to conclude that later stages are built on the prior ones. The method of minimal anatomies will tell us when new, more advanced neural network structures become necessary, and Critical semantics will provide a theoretical foundation for figuring out why these new structures were not evident in the previous stages. The short answer to this will be: because in their undeveloped state these structures did not form locally path connected (l.p.c.) trajectories in a semantic topological space [Wells (2011c)], which means they failed to produce meaningful representations in *nous* and, therefore, could not be *systematically expressed* as circular reactions and learned habits. It isn't much of a conjecture to make that we will find something new appearing in Martian 2 behavior when they are added, namely, the seemingly-random movements and expressions that are so characteristic of newborn infant behavior.

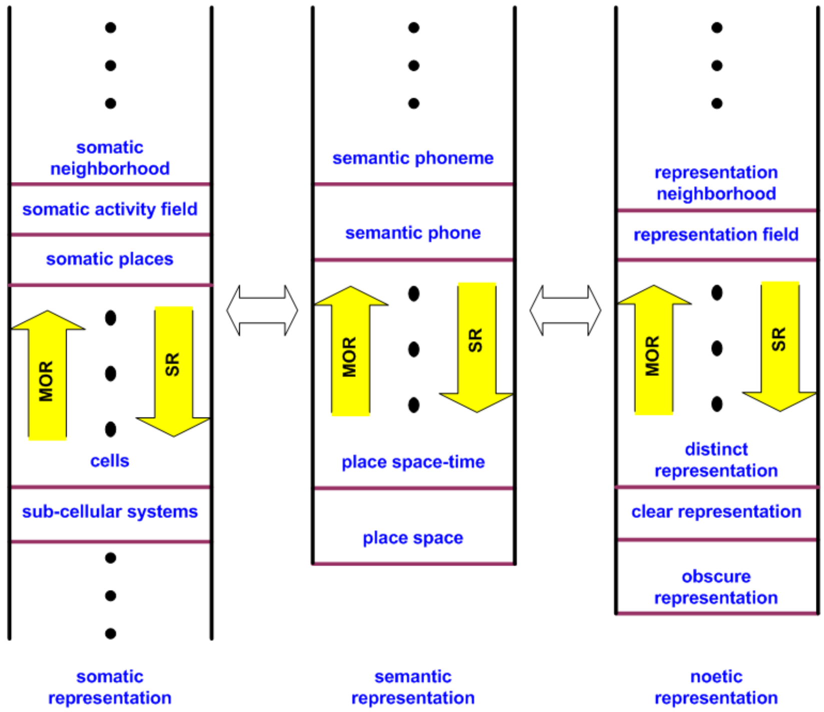

Figure 7: Somatic-semantic-noetic ladder structures as a tactical schema for psychophysical research.

This tactical approach is made systematic by the development of the applied metaphysic of the somatic code [Wells (2011a)]. Prior to this point, stage-by-stage modeling of sensorimotor development promised to be a more or less heuristic affair. The Critical semantics of the somatic code has now provided us with what I call a ladder structure for scientific reduction (SR) and for model order reduction (MOR). Figure 7 illustrates this concept, which was introduced in more detail in [Wells (2011c)]. As much as I would like to take full credit for the ladder, in fact I can do no more than claim the formal Critical semantics portion of the idea. The rest of it comes from Grossberg's earliest works in embedding field theory, particularly Grossberg (1969). Forty-two years after the guy who broke the tape is too far back in the marathon pack for a runner to claim any medals.

We were able to get Martian 1 to exhibit perception-stimulated practice of innate sensorimotor schemes very quickly, and we were even able to demonstrate a *learned* aversion response to an unpleasant stimulus ("pain") [Wells and MacPherson (2009)]. However, we were much less successful in getting Martian 1 to move from reflex actions to cognitively learned habits (which marks the beginning of stage 2 of sensorimotor intelligence). Analysis of its shortcomings shows that they come from two factors. The first, and probably simplest to overcome, was that we did not include a "temporary memory" structure – by which I mean the sort of memory phenomenon we are all familiar with from, e.g., looking up a phone number and not being able to remember it the second we've finished dialing the phone. There are many possible technical ways to address this. Grossberg has discussed several over the years. One of the simplest and most direct ways, at least in principle, is to use ring avalanche networks since the function requires learning plasticity but also learning instability [Wells and MacPherson (2009)]. We had some tantalizing signs of success with this on Martian 1 just before that program had to be terminated.

But the major deficiency in Martian 1 in regard to acquiring learned habits was due to a lack of adequate *affectivity* in the very primitive "limbic system proxy" it employed. This deficiency was a direct consequence of basing Martian 1 on Brooks' model rather than on figure 1. Seen in hindsight, it was a foolish mistake. Although in adults cognition and affectivity contribute more or less equally to behavior, in the infant this is not the case. Infants lack a sufficient basis of experience-based cognitions in the manifold of concepts (a structure within determining judgment in figure 1) for cognition to have much behavioral effect. Rather, the infant's very earliest cognitive growth is driven by *affective* preferences. The empirical evidence of this is very compelling. Greenspan's work with severely autistic children [Greenspan (1979, 1997)] directly supports the theory. Piaget, too, documented evidence for stage-by-stage affectivity development proceeding in parallel and in interaction with cognitive development [Piaget (1954)]. Further evidence can also be seen in reports detailing therapy treatments of boys suffering severe neuroses as a result of sexual abuse [Nyman and Svensson (1995), esp. the case of "Patrik," chap. 1, 11]. However, the most compelling factor here comes from mental physics, which tells us *as a theorem* that early infantile cognition is affectively driven (process of reflective judgment) *necessarily*. One of the earliest projects in the Martian 2 program is to replace the Martian 1 "limbic system proxy" with a semantic-topology-based neural network anatomy.

On the whole, psychology has not dealt well with "emotion" and "motivation" questions. Everyone admits that some "emotional-motivational system" must exist somewhere in brain structure, and the limbic system is regarded by most as a major part of it. Yet psychology has been curiously slow to "warm up" to emotion psychology even though emotion theories go back to James-Lange and "motivational psychology" was introduced early in the twentieth century. On the whole, emotion and motivation theories and models have not been successful. This is discussed at length in chapter 15 of Wells (2006) as well as in Reber's *Dictionary*.

It is not difficult to understand why this is. The theory of mental physics is very specific on this point. The major mini-theories – of which a list of some of the better representative efforts would include Plutchik (1980), Buck (1988), Russell (1997), and many others reviewed in Carlson and Hatfield (1992) and Lewis and Haviland-Jones (2000) – fail because they begin with *ad hoc* definitions of "emotion" or "motivation" that lack objectively valid grounds and ignore the fundamentally epistemological nature of affectivity phenomena.

Of the current mini-theories, Russell's "script theory" comes closest in mathematical form to a set membership treatment. His method (though not his theory) might serve as an example for ideas of a formal SMT approach to affectivity, but in point of fact we need to begin with the five synthetic process of sensibility, the *momenta* of reflective judgment and an understanding of the motivational dynamic in Wells (2009). From there the next step is to develop the semantic topology structuring, and go on from there to embedding field synthesis. Embedding fields for the synthesis of apprehension & apperception and for the process of reflective judgment are the two highest priority topics for the initial Martian 2 research efforts.

VI. References

Abeles, Moshe (1991), *Corticonics*, Cambridge, UK: Cambridge University Press.

Brennan, Aaron (1998), *Binary Connectionist Networks*, M.S. Thesis, Moscow, ID: The University of Idaho.

Brooks, Vernon B. (1986), *The Neural Basis of Motor Control*, Oxford, UK: Oxford University Press.

Buck, Ross (1988), *Human Motivation and Emotion*, 2nd ed., NY: John-Wiley and Sons.

Carlson, John G. and Elaine Hatfield (1992), *Psychology of Emotion*, Ft. Worth, TX: Harcourt Brace Jovanovich College Publishers.

Carpenter, Gail A. (1980), "Normal and abnormal signal patterns in nerve cells," *SIAM AMS*

Proceedings: Mathematical Psychology and Psychophysiology, Providence, RI: American Mathematical Society vol. 13, pp. 49-90.

Chen, C. T. (1970), *Introduction to Linear System Theory*, NY: Holt, Rinehart and Winston.

Combettes, Patrick L. (1993), "The foundations of set theoretic estimation," *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182-208.

Greenspan, Stanley I. (1979), *Intelligence and Adaptation*, NY: International Universities Press.

Greenspan, Stanley I. (1997), *The Growth of the Mind*, Reading, MA: Addison-Wesley.

Grossberg, Stephen (1968), "Some physiological and biochemical consequences of psychological postulates," *Proceedings of the National Academy of Science*, 60: 758-765.

Grossberg, Stephen (1969), "Embedding fields: A theory of learning with physiological implications," *Journal of Mathematical Psychology*, **6**, 209-239.

Grossberg, Stephen (1971), "Embedding fields: Underlying philosophy, mathematics, and applications to psychology, physiology, and anatomy," *Journal of Cybernetics*, vol. 1, 28-50.

Grossberg, Stephen (1972), "A neural theory of punishment and avoidance, I: Qualitative theory," *Mathematical Biosciences* **15**, 39-67.

Grossberg, Stephen (1973), "Contour enhancement, short term memory, and constancies in reverberating neural networks," *Studies in Applied Mathematics*, vol. 52, no. 1, pp. 213-257.

Grossberg, Stephen (1975), "A neural model of attention, reinforcement and discrimination learning," *International Review of Neurobiology*, vol. 18, pp. 263-327.

Grossberg, Stephen (1980), "Psychophysiological substrates of schedule interactions and behavioral contrasts," *SIAM AMS Proceedings: Mathematical Psychology and Psycho-physiology*, Providence, RI: American Mathematical Society vol. 13, pp. 157-186.

Hall, William C. and Adonis Moschovakis (2004), *The Superior Colliculus*, Boca Raton, FL: The CRC Press.

Kailath, Thomas (1980), *Linear Systems*, Englewood Cliffs, NJ: Prentice-Hall.

Lewis, Michael and Jeannette M. Haviland-Jones (2000), *Handbook of Emotions*, 2nd ed., NY: Guilford Press.

McCarthy, Steven G. and Richard B. Wells (1997), "Model order reduction for optimal bounding ellipsoid channel models," *IEEE Transactions on Magnetics*, vol. 33, no. 4, pp. 2552-2568.

Milanese, M. and A. Vicino (1991), "Optimal estimation theory for dynamic systems with set membership uncertainty: An overview," *Automatica*, vol. 27, no. 6, pp. 997-1009.

Nelson, David (2003), *Dictionary of Mathematics*, 3rd ed., London: Penguin Books.

Neumann, John von (1948), "The general and logical theory of automata," in *Papers of John von Neumann on Computing and Computing Theory*, William Aspray and Arthur Burks (eds.), Cambridge, MA: The MIT Press, 1987, chap. 10, pp. 391-431.

Nyman, Anders and Börje Svensson (1995), *Boys Sexual Abuse and Treatment*, London: Jessica Kingsley Publishers, 1997.

Piaget, Jean (1954), *Intelligence and Affectivity: Their Relationship During Child Development*, Palo Alto, CA: Annual Reviews, Inc., 1981.

Piaget, Jean (1970), *The Principles of Genetic Epistemology*, NY: Basic Books, 1972.

Piaget, Jean and Rolando Garcia (1987), *Toward A Logic of Meanings*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

Plutchik, Robert (1980), *Emotion: A Psychoevolutionary Synthesis*, NY: Harper & Row.

Rao, A. and Y.-F. Huang (1993), "Tracking characteristics of an OBE parameter estimation algorithm," *IEEE Transactions on Signal Processing*, vol. 41, pp. 1140-1148.

Reber, Arthur S. and Emily S. Reber (2001), *Dictionary of Psychology*, 3rd ed., London: Penguin Books.

Russell, James A. (1997), "How shall an emotion be called?" in *Circumplex Models of Personality and Emotions*, Robert Plutchik and Hope R. Conte (eds.), Washington, D.C.: American Psychological Association, pp. 204-220.

Schetzen, Martin (1980), *The Volterra and Wiener Theories of Nonlinear Systems*, NY: John Wiley and Sons.

Sherman, S. Murray and R. W. Guillery (2006), *Exploring the Thalamus and Its Role in Cortical Function*, 2nd ed., Cambridge, MA: The MIT Press.

Walter, Eric and Hélène Piet-Lahanier (1990), "Estimation of parameter bounds from bounded-error data: A survey," *Mathematics and Computers in Simulation*, 32, pp. 449-468.

Wells, Richard B. (2011a), "The applied metaphysic of the somatic code," June 16, available on the author's web site.

Wells, Richard B. (2011b), "On Critical representation in brain theory, Part I: Critique," May 25, available on the author's web site.

Wells, Richard B. (2011c), "On Critical representation in brain theory, Part II: General schema of knowledge representation," June 30, available on the author's web site.

Wells, Richard B. (2011d), "On the derivation of an applied metaphysic," May 20, available on the author's web site.

Wells, Richard B. (2011e), "On Critical doctrine of method in brain-theory," March 31, available on the author's web site.

Wells, Richard B. (2011f), "Weaver's model of communication and its implications," June 2, available on the author's web site.

Wells, Richard B. (2011g), *Unabridged Glossary of the Critical Philosophy and Mental Physics*, available on the author's web site.

Wells, Richard B. (2010), *Leadership*, available on the author's web site.

Wells, Richard B. (2009), *The Principles of Mental Physics*, available on the author's web site.

Wells, Richard B. (2006), *The Critical Philosophy and the Phenomenon of Mind*, available on the author's web site.

Wells, Richard B. and Quinn MacPherson (2009), "The Martian Program 2009: Theory of the stage I infant model," April 20, 2009, available on the author's web site.

Winer, Jeffery A. and Christoph E. Schreiner (2005), *The Inferior Colliculus*, NY: Springer.