Computational Synthesis: Following the Treaded Mathematical Track

Zippora Arzi-Gonczarowski
Typographics, Ltd. Jerusalem 96222 Israel
zippie@actcom.co.il
http://www.actcom.co.il/typographics/zippie

Abstract

A mathematical approach to intelligent agent modeling is described, where one starts from low level sensory-reactive mechanisms and proceeds, by mathematical synthesis, to scale to high complexity and to model high level, domain independent, intelligent functionalities. The composition of building blocks features regularity, modularity, abstraction, and awareness of hierarchy.

The General Approach

The essence of mathematical modeling has always been to start from basic, low-level, building blocks which are intuitively convincing and obvious. Then, following a long series of simple steps, to obtain arbitrary high-level constructs. The typical paradigm is the system of natural numbers: The five postulates of Peano capture the pre-theoretical essence of the natural numbers as counters of discrete quantities. Orderly extensions of the natural numbers provide the integers, then the rational numbers, then the real numbers. One starts from fundamental concepts as primitive terms, and asserts certain simple propositions (postulates, axioms) about them. Further terms are then introduced in an orderly manner, using the primitive terms. Theorems express properties of these new terms, applying deductive reasoning to obtain them from the postulates. Another typical paradigm is Euclidean geometry, that models physical space using points and lines as building blocks. 20th century 'pure' mathematicians, considering such theories, perceived that they shared domain independent principles and methods (Lawvere & Schanuel 1997). Not the least ones of these abstracted principles and methods are about the identification and the composition of low-level building blocks to scale to high-level constructs, achieving arbitrary complexities.

Hilbert coined the term *The Genetic Method* for the method which is suggestive in the contexts of AI and of biological complexity: Natural intelligent systems started evolving from the earliest nerve cell that was probably a combined receptor (receiving environmental stimuli) and motor unit (producing muscle or gland response). With biological systems as role models, intelligent systems could be modeled mathematically by starting from primitive building blocks that capture an abstraction of that, and orderly structured extensions could then be introduced to model higher level funcionalities, applying deduction to obain and to study their properties. Reusage of sub-systems is frequent in the biological domain as well: Evolution theorists use the term *exaptations* (Gould & Vrba 1982) to refer to minor changes that make use of already existing capabilities to create new behaviors. Exaptation can be readily modeled mathematically by structural abstraction and re-application of relevant constructs and proofs.

One 'no free lunch' price for generic domain independent models seems to be an extreme care to balance between abstraction that is not detached, and grounding that is not over deterministic. In particular: $(\imath)$ A bimodality of mathematical modeling is that, on one hand, one treats the concepts *as if* they were meaningless. That should warrant that all assumptions are stated explicitly as postulates, and no hidden properties of the primitive terms enter from their pre-theoretical, commonsense, intuitions. On the other hand, one needs to invariably refer to the domain that is being modeled, making certain that the theory that emerges *validly* describes the phenomena that gave rise to the formalization. $(\imath\imath)$ In AI, the open-ended diversity of phenomena that need to be modeled has made it difficult to capture things in a uniform manner and to benefit from mathematization as a powerful modeling tool like other scientific domains. If one were to overcome this last obstacle, then this would probably be by finding a *higher, yet suitable,* level of abstraction: high enough to absorb a variety of phenomena, but not so high as to conflate all meanings.

Category theory (Herrlich & Strecker 1973; MacLane 1972) has been developed precisely for such purposes within mathematics itself. It provides meticulous tools of rigour to capture a structural essence without being over deterministic. Category theory has already been successfully applied to various issues in computer science, like programming language semantics and the design of programs using abstract data types, providing a standard ontology and language of discourse for these areas of research (Barr & Wells 1995). Its advantages for modeling intelligence have been solicited, among others, by (Magnan & Reyes 1994).

ISAAC, an Integrated Schema for Affective Artificial Cognition, (Arzi-Gonczarowski & Lehmann 1998b; 1998a;

Arzi-Gonczarowski 1998; 1999b; 1999a; 2000b; 2000a; 2001a; 2001b; 2002) is a formalism that follows the above guidelines. It boots agents' 'minds' from a formalization of reaction driven perceptions, abstracting an essence of the natural evolutionary context. It then scales up to complex high level functions, using a rigorous mathematical framework and a circumbscribed number of reusable and composable concepts and constructs, thus yielding a continuum from low- to high- level intelligence. The scope of this paper limits us to an intuitive, hand waving, synopsis of mathematical results. Readers interested in theorem proofs and technical detail are encouraged to refer to the longer articles referenced above, available at the author's web page. They also ground the schema in context specific examples.

ISAAC's Basic Building Blocks

One effective tradition of foundational scientific research has been to go back to first principles in order to grapple with an issue. If intelligence is the end, then what are the first principles of intelligence? (Allen 1998) says: *'a prerequisite for something to be intelligent is that it has some way of sensing the environment and then selecting and performing actions.'* If intelligence boils down to a sensible marriage between behavior and circumstances as a first principle, then the building blocks should be about that. 'Sensible' in this context would be relative to agents' concerns: survival, and the pursuit of various goals (of course, the agent is not necessarily 'aware of its concerns'). Behaviors are more than often conjured up as responses to stimuli in the environment, hence agents are provided with a sensing apparatus, and the building blocks should account for that as well.

In the biological context, evolution naturally selected sensory motor neural apparatuses that coupled embodiments of organisms with their ecological niches, yielding behavior designated as 'intelligent' because it happened to support endurance of the species. In the artificial context agents are typically constructed to serve a purpose, so that 'intelligent' behavior is goal-directed. However, survival is often a concern in that context as well: The setting of agents in external environments exposes them to hazards that could not always be expected. Material existences in real physical environments as well as virtual entities in 'cyber' environments are in jeopardy. They can be injured and incapacitated. In dynamic environments some of the protective measures should be typically reactive: agents should be able to sense danger as it comes and to react, often urgently, in an appropriate manner to safeguard their existence. In both natural and artificial contexts, sensations and reactions should be tightly coupled, as they determine each other: Suitable reactions are conjured up by discriminating sensations that are, in turn, tailored for the forms of behavior that are afforded.

It has long been accepted that forms of natural intelligence, including sublime ones, are results of (combinations of) improvements and upgrades applied, by natural selection, on top of basic reactive intelligence. Evolution upgraded intelligent agents from simple reactive organisms in a random and cluttered manner, patch over patch. Though the results are a tantalizing living proof that scruffy design could work, one might try an orderly approach when given a chance to consciously and systematically design artificial intelligences, keeping a neater record of that which goes on.

The proposed mathematical model starts from basic building blocks that stand for basic discriminating sensations and basic reactions that go with them. In the service of domain independence and of mathematical abstraction, the specific nature of these building blocks is left undetermined: an agent's perception $\mathcal{P}$ makes a discrimination α about some world chunk w, and that conjures up a reaction r (a formal definition follows below). Substitution instances of the schema should provide these deliberately meaningless symbols a concrete substantiality.

Based on these building blocks, and following a series of mathematical steps, the proposed theory obtains high-level cognitive, behavioral, and affective constructs and functions of intelligence. Abstraction of construct functionality and reusage of structures occur naturally along the way, providing us with positive feedback that the proposed premises are probably useful and adequate.

Basic Objects

All the papers cited at the end of the first section share the following premises, with further extensive discussions of the methodical considerations behind these premises[1].

Definition: A *Perception* is a 5-tuple $\mathcal{P} = \langle \mathcal{E}, \mathcal{I}, \varrho, \mathcal{R}, \mathcal{Z} \rangle$

- $\mathcal{E}, \mathcal{I}, \mathcal{Z}$ are finite, disjoint sets
- ϱ is a 3-valued predicate $\varrho : \mathcal{E} \times \mathcal{I} \rightarrow \{t, f, u\}$.
- $\mathcal{R}$ is a function: $\mathcal{R} : \mathcal{I} \longrightarrow \mathcal{Z}$

The set $\mathcal{E}$ stands for a 'snapshot' of a perceived environment: its elements model environmental chunks, most typically objects or events, (*world elements w-elements*) that could perhaps be discerned by an agent. Even if the environment exists independent of its perception, its carving up into individuated w-elements typically depends on the agent: One perceives a forest where another perceives many trees. Later, the concept of perceived environments will be abstracted, to be reused also for conceived environments that are just imagined, or recalled, thus reusing and scaling the same definition to model higher functions of intelligence.

The elements of $\mathcal{I}$ model discriminations that are afforded by the perceiving agent, named *connotations*. Each connotation models a 'socket' where a discriminated stimulus should 'plug in' (the 'plugging' is modeled by ϱ as described below). For example, if the environment of a perception features a w-element ALARM_BELL that happens to be sounding, perception could 'plug it' into the connotation *alarm_sound* (and possibly also to other connotations, such as *loud_sound* or *interrupted_sound*). When reactions are 'wired' to that connotation, they would be then conjured up (that is modeled by $\mathcal{R}$, described below). Later, the role of connotations will be extended beyond mere 'sensation sockets', to labeled representations that stand for the relevant stimuli. They can then be accessed and summoned by

[1]In the course of a few years of research the notation and the terminology underwent a few adjustments and extensions, that do not effect the essence and the applicability of earlier results, which are still effective and provide a basis to the proposed theory.

internal 'thought' processes. The same definition will thus be reused and scaled to model representational capabilities of higher intelligence.

The 3-valued *Perception Predicate* (*p-predicate*), ϱ, models perception of connotations in w-elements: $\varrho(w, \alpha) = t$ stands for definitely 'yes', namely 'w has the connotation α'. $\varrho(w, \alpha) = f$ stands for definitely 'no', namely 'w does not have the connotation α'. $\varrho(w, \alpha) = u$ indicates that perception, for some reason, does not tell whether the stimulus for which α stands, is detected in w or not. In the example above, if the agent does perceive that the bell sounds, one gets $\varrho(w, \mathit{alarm\_sound}) = t$, If the agent perceives that the bell is quiet then one gets $\varrho(w, \mathit{alarm\_sound}) = f$. $\varrho(w, \mathit{alarm\_sound}) = u$, if, for instance, the environment is too noisy to tell, or the agent's hearing is impaired.

The elements of $\mathcal{Z}$ stand for procedures, or methods, that the relevant perception can activate, modeling a set of behaviors for that perception. For basic reactive agents, it would typically consist of that agent's set of physical reactions (`feed`, `fight`, or `flight` are common examples). If `null` is an element of $\mathcal{Z}$, then it stands for the empty call, namely no response, or indifference. Later, the concept of behavior will be extended to introvert mental behavior, to innately motivated behavior that is not necessarily triggered by outside stimuli, and to behavior that could be just conceived without being consummated. The same definition is thus scaled and reused to model complex, sophisticated, forms of intelligence.

$\mathcal{R}$ models reaction activation. If, for some w and α, $\varrho(w, \alpha) = t$, then $\mathcal{R}(\alpha)$ is conjured. For example: $\mathcal{R}(\mathit{alarm\_sound}) =$ `flight` models an agent that runs away when it hears an alarm sound. Procedures, such as `flight`, would be simple and deterministic for low level systems. Later, this framework reuses the same definition to model upscaled forms of intelligence, by extending the concept of reaction to include control over other reactions, more complex procedures, and to activations that are triggered by the *absence* of stimuli.

The proposed formalization of w-elements and connotations does not necessarily mean that w-elements need to be always elaborately analyzed. To make room for non-analytic discriminations of (and reactions to) synthetic wholes, perceptions could have w-elements w in the environment associated with connotations of the form $''w''$, that provide holistic discriminations.

Specific instantiations of the five coordinates of perception: $\mathcal{E}$, $\mathcal{I}$, ϱ, $\mathcal{Z}$, and $\mathcal{R}$, provide specific perceptions. The mathematical objects $\mathcal{P}$ hence stand for basic embodied reactive precognitions. They are high-level in the sense that they are presumed to layer on top, and be grounded by, a sensory motor neural apparatus. They are low level (in another sense) if these coordinates never change. In that simplistic case our story ends here: These perceptions could then be easily programmed using a loop that checks the sensors and reacts accordingly, practically conflating ϱ with $\mathcal{R}$.

Basic Dynamics of Objects

Agents modeled by static building blocks from the last subsection cannot even handle simple environmental changes, let alone refine or generalize discriminations, or adapt their behavior. Indeed, in the natural context, if a low level organism is moved into an ecological niche which is different from its natural habitat, it is unlikely to cope with the new setting and to survive. If it does manage to adjust and to survive, then we would tend to say that it somehow has 'a higher degree of intelligence'. The evolutionary pressure to afford change is hence clear. To model that, one needs tools to describe transitions from one perception into another. Flows of conversion between perceptions are formalized by *perception morphisms (p-morphisms, arrows)*:

Definition: Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two perceptions:

$$\mathcal{P}_1 = \langle \mathcal{E}_1, \mathcal{I}_1, \varrho_1, \mathcal{Z}_1, \mathcal{R}_1 \rangle \ , \ \mathcal{P}_2 = \langle \mathcal{E}_2, \mathcal{I}_2, \varrho_2, \mathcal{Z}_2, \mathcal{R}_2 \rangle$$

A p-morphism $h : \mathcal{P}_1 \rightarrow \mathcal{P}_2$ is defined by the set mappings:

$$h : \mathcal{E}_1 \rightarrow \mathcal{E}_2 \quad , \quad h : \mathcal{I}_1 \rightarrow \mathcal{I}_2 \quad , \quad h : \mathcal{Z}_1 \rightarrow \mathcal{Z}_2$$

With the following structure preservation conditions: $(\imath)$ **No–Blur:** For all w in $\mathcal{E}$, and for all α in $\mathcal{I}$, If $\varrho_1(w, \alpha) \neq u$, then $\varrho_2(h(w), h(\alpha)) = \varrho_1(w, \alpha)$. $(\imath\imath)$ **Disposition:** For all α in $\mathcal{I}$, If $\mathcal{R}_1(\alpha)) \neq$ `null`, then $\mathcal{R}_2(h(\alpha)) = h(\mathcal{R}_1(\alpha))$.

The rest of this subsection discusses issues of this definition (with the reservation that, had verbal descriptions been able to grasp the full implication of mathematical definitions with total clarity and precision, then one may have done without the mathematization in the first place).

The set maps are formal tools to describe transitions in the perceived environment, or changes in discriminations, or changes in behavior, respectively. Indeed, the mathematics of set maps affords distincitions between 'onto' versus not 'onto', and 'one-to-one' versus 'many-to-one'. A few examples: in a set map $\mathcal{E}_1 \rightarrow \mathcal{E}_2$ that is not 'onto', $\mathcal{E}_2$ may feature a new w-element, say another agent w, that is not part of $\mathcal{E}_1$ (maybe it just arrived). In a set map $\mathcal{I}_1 \rightarrow \mathcal{I}_2$ that is not 'onto', $\mathcal{I}_2$ may feature a new connotation, say *fragile*, indicating the learning of a new discrimination. Likewise, $\mathcal{Z}_2$ may feature a new behavior, say `alarm_turn_off`. In set maps that are not 'one-to-one', constituents (w-elements, connotations, behaviors) may be merged, modeling amalgamations, generalizations, combinations, etc. Set maps also enable conversion of one constituent into another. Replaced connotations typically indicate translations or interpretations of discriminations, replaced w-elements typically indicate similarity of objects in some respect, (such as in analogies (Arzi-Gonczarowski 1999b)), and replaced behaviors typically indicate adaptation of reactions.

P-morphisms have been shown to be flexible enough to formalize and to model a broad spectrum of cognitive, affective, and behavioral flow, from small alterations (that slightly update only few constituents) to transformations so profound that $\mathcal{P}_1$ and $\mathcal{P}_2$ may appear to have little in common. The structure preservation conditions on p-morphisms are related to the notoriously evasive core invariable aspect of meaning that one would like to preserve, even if loosely, across contexts. Keeping track of a sensible process of perceptual-cognitive change is formalized here by the structure preservation *No-Blur* condition: Values of the p-predicate may be modified along arrows, but that is confined by the *No-Blur* condition, which binds change in the environment $\mathcal{E}$ with change in the interpretation $\mathcal{I}$. Transitions between w-elements need to be justified by commensurate connotations, and, on the other hand, transitions between connotations need to be grounded by commensurate experience. A *factorization theorem* in (Arzi-Gonczarowski 1999b) shows how to keep a rigorous track of even the most complex transitions.

Keeping track of a sensible process of behavioral flow is formalized by the structure preservation *Disposition* condition: Behaviors and reactions may be modified along arrows, but that is confined by that condition, which binds change in interpretation with change in behavior. Specific contexts may, of course, add their own structure preservations.

Mathematical Framework

Technically, composition and the identity p-morphism are defined by composition and identity of set mappings, and it has been shown that perceptions with p-morphisms make a mathematical category, designated $\mathcal{P}rc$. Category theory provides a well developed mathematical infrastructure to capture the structural essence of perceptive behavior and intelligent processes, without being over deterministic. The basics of the proposed category are presented in (Arzi-Gonczarowski & Lehmann 1998b). Results are invariably inferred and concluded only from the formal premises using mathematical tools and methods. However, whenever a result is reached, it is vital to examine it with regard to the pre-theoretical considerations, and to test it against existing theories and opinions about intelligent systems.

It is sometimes helpful to consider a category, and $\mathcal{P}rc$ in particular, as a graph. In that underlying structure, perceptions are vertices and p-morphism arrows are edges. Formalized intelligence processes perform by construction and navigation of relevant portions of that graph. Transitions between perceptions in this setting are modeled by edge paths in the graph. Complications to this simplification typically arise from more than one way of getting from one vertex to another, which is often the result of compositions of arrows, and their counterpart - factorizations of arrows. That is where *theorems about commutative diagrams* come into the picture, stating when one path is equivalent to an alternative one. In the proposed category theoretical setting, theorems about commutative diagrams are the theoretical results. (Barr & Wells 1995, p.83) entitles commutative diagrams as *the categorist's way of expressing equations*.

Computational Synthesis, So Far

An obvious pressure for the introduction of p-morphisms is probably the need for descriptions of changes that occur with *time*, and the arrow then coincides with the arrow of time. However, at the pure formal level, an arrow just models a structural commensuration of two perceptions. This abstraction opens the possibility to apply p-morphisms to model other types of transitions and relationships, where the arrowed representation is not necessarily chronological[2]. A target perception of some p-morphism could, for example, exist prior to the domain perception of that p-morphism. In that case, from the chronological point of view, the arrow is transitioned 'backwards', modeling perceptual values that are being blurred, constituents that are being deleted, discriminations that are being refined, and so on.

Another useful application of p-morphisms is inter-agent, rather than intra-agent. Inter-relating between different agents' perceptions provides basis for modeling paths of communication. In that case, both the domain and the target perception of a p-morphism would exist first, and arrows would then be constructed to bridge between them (ISAAC affords formal algorithmic procedures for that purpose.)

Generally speaking, a regularity of all functionalities and modules in ISAAC is that they consist of (structured compositions of) p-morphisms: Interpretive transitions, representation formations, analogy making, creative design, intra-agent communications, joint perceptions, and more.

Having introduced arrows between perceptions, one can immediately integrate this building block into prior framework. In the basic definition of perceptions, the set $\mathcal{Z}$ models a collection of behaviors for that perception. The activation of a p-morphism could be a legitimate behavior, too, thus extending the notion of behavior to include this type of transitions. No additional definitions are needed for that. The function $\mathcal{R}$ could now have a value: $\mathcal{R}(\alpha) = \text{ACTIVATE}(h)$, modeling an agent that changes its state in response to some perceptual discrimination.

The implications of this last composition of building blocks could be far reaching, and it has the potential of scaling the system to surprising complexities. Since perceptions, as defined, determine reactions, a transition $h : \mathcal{P}_1 \to \mathcal{P}_2$ may involve a change in (some) reactions. As an example, consider an agent that perceives how the environment responds to one of its reactions, and is hence impelled to undergo a transition h to a modified state with that behavior toned up (reinforcement) or down, according to the perceived response. One intriguing property of this combination is that the activation of h is not necessarily overt. The change would be eventually observed from the outside only when a relevant overt reaction is at all conjured, which may happen after a long delay, when the external catalyst that caused the transition is no longer there. Figuring out the course of change would be somewhat like psychoanalysis.

The general idea is that intelligent systems should be initialized to 'genetic' perceptions, featuring mininal basic essential constituents as bootstrap. (All arrows are bound on their left by the *empty perception*, the initial object of the category, modeling a theoretical *Tabula rasa*.) If a system is at all capable of p-morphisms, then no additional definitions are required to inspirit the system to mature. Perceptual transitions would be triggered, uniformly like everything else, by the reaction function $\mathcal{R}$, either $(\imath)$ systematically, by activation of an arrow $\mathcal{R}(\alpha) = \text{ACTIVATE}(h)$, as just explained, or $(\imath\imath)$ by a leap: $\mathcal{R}(\alpha) = \texttt{leap\_to}(\mathcal{P}')$. The first option models a transition that could be analyzed by p-morphisms, while the latter option opens the possibility to model, if so desired, wilder 'mental jumps', that are sometimes entitled *Proust effect*: a stimulus 'throws' the agent to an altogether

[2] This is an example instance where insisting on a chronological interpretation of arrows would be introducing properties, that are not formally there, from some pre-theoretical intuition.

different mental state.

P-morphisms are themselves made of regular modules: (Arzi-Gonczarowski 1999b) parallels the $\mathcal{I}$ mapping ($h : \mathcal{I}_1 \to \mathcal{I}_2$) of a p-morphism as the *interpretive* component of the transition, with the $\mathcal{E}$ mapping ($h : \mathcal{E}_1 \to \mathcal{E}_2$) of the same p-morphism, as the *literal-analogical* component of the transition. The $\mathcal{E}$ mapping is 'pro-synthetic' in that it takes cohesive, existing, w-elements as its basic building blocks and maps between them. The $\mathcal{I}$ mapping is 'pro-analytic' in that it 'breaks' impressions of cohesive whole into particular discriminations as building blocks, and maps between them. Computationally abstracted, both are set maps. A salient property of the basic definition is the symmetry between w-elements and connotations as variables of the p-predicate. From a purely technical, context free, point of view, their roles are interchangeable. This *duality* has theoretical and computational consequences (Arzi-Gonczarowski 1999a). For example, any formal construction or theorem that is established for connotations (w-elements) can automatically be applied to w-elements (connotations), mutatis mutandis. This suggests insights into a 'connaturality' of cognitive processes and capabilities and, as basis for computational implementations, this entails architectural and applicational modularities.

Scaling to Higher Complexities

Except for lowest level contexts, any non-trivial perception would either have an environment of more than a single w-element, or would have to deal with more than a single stimulus, and hence more than a single reaction at a time. Conflicting behaviors would be conjured, that could not be performed simultaneously, bringing about confusion and disordered behavior. That constitutes a natural pressure (not the only one) to handle combinations of constituents in an orderly manner, and to model that.

Already at an intuitive level, Boolean combinations of elements (using *and, or,* and *not*) seem to provide an exhaustive collection of possible combinations. That was, perhaps, the intuition that guided George Boole when he introduced Boolean algebra in his 1854 statement (Boole 1854) *An Investigation of the Laws of Thought*. In the context of ISAAC, it has been shown that closing a set of constituents (such as the connotations $\mathcal{I}$, or w-elements $\mathcal{E}$) under Boolean connectives, provides infrastructure for achieving a wide range of high-level intelligence functionalities. Based on results of the well developed theories of Boolean algebras and of categories, an automatic and methodical closure of basic constituents into Boolean lattices is formulated. ISAAC applies this abstract procedure to the various constituents of perception. (The p-morphism tool is upscaled accordingly by letting relevant set maps be Boolean homomorphisms.)

Lattices of Connotations and Related Reactions

In the setting of the basic definition, whenever perception discriminates in its environment a w-element with a certain connotation (as defined by ϱ), perception 'plugs it' into the 'socket' associated with that connotation, triggering reactions that are 'wired' to that 'socket' (as defined by $\mathcal{R}$). At that basic level, if more than one connotation is perceived and relevant reactions triggered, *one would not 'know' about the other*, with no coordination between them. Confusion and disordered behavior could easily follow. Assume now that all these connotations are interconnected and arranged in a lattice, where every combination of connotations corresponds to a junction node in the lattice. The basic reactions are innate, and are hence invariably stimulated *because an emergency could be involved*, but, at the same time, there is also referral to the relevant combination node. At that node one may develop mechanisms that are designed to arbitrate and to salvage confusions that could be under way.

The proposed lattice just provides infrastructure where arbitrating mechanisms could be wired. Specific arbitration solutions would be a domain specific issue. At a simple level, those could consist, for example, of a mechanism of automatic prioritization and selection: one selected reaction is consummated, and the conflicting ones are suppressed. A higher level option would be to creatively substitute, or to integrate, essential elements from a few behaviors into one coherent behavior that perhaps compromises a little, but takes care of almost everything. It is easy to see that solutions to these conflicts would often involve a suppression of some basic reactions that have already been stimulated. The basic innate reactions are typically about vital concerns, and hence they are likely to be vigorous and perseverant. In that case, a period of dissolution is expected, while these suppressed impulses persist as they are fading out, and energy is being invested in containing that process. (Arzi-Gonczarowski 2002) suggests similarities between that and human emotions. Following (Frijda 1986), who defines the core of an emotion as the readiness to act in a certain way, the reaction function $\mathcal{R}$ is hence upscaled to also include an emoting aspect, as well as a control aspect, in higher level intelligence. For lower level systems it models reactions that are invariably performed, while for higher level ones it could also model action tendencies that are not consummated.

The ensuing engineering perspective of intelligent behavior is essentially about management, maintenance, and amelioration of a large household of adamant action tendencies. The Boolean closure introduces higher order ones, so that the system's behavior, that is finally and actually generated, should be sensible. A significant design principle is about hierarchy: one is not allowed to deny the legitimacy, or get rid, of the lower level, innate, action tendencies. One is only allowed to toy with smarter, and more adamant, controllers, arbitrators, diverters, negotiators, reasoners, and so on.

A behavioral pressure for the introduction of lattices of connotations has just been described. This is now followed by showing how these lattices can be reused to serve other significant interests of intelligence. In a natural evolution cotext one might have said that they *exapted* to be the *Laws of Thought*. Memory, anticipation, planning and reasoning are all intelligent capabilities and skills that developed in order to better understand, and thus to prepare for, various things that could happen in the environment. They all require, first of all, an internal representational apparatus. To scale up for that, all one needs is *to be able to refer* to the building blocks that already represent discriminations: the connotations. When these building blocks are labeled, they are able to span the possible content of a representation and the ontological distinctions that can eventually be made.

The following features of complemented and distributive lattices, namely Boolean algebras (Sikorski 1964), of labeled connotations, serve representational purposes and related procedural objectives: (A) They feature a partial order. This may enable the organization of connotations in *taxonomic hierarchies*, with inheritance of information. (B) They feature the two binary operations $\vee$ and $\wedge$, and the unary operation $\neg$, allowing the formation of *compound concepts* as combinations of more basic concepts [3]. (C) The lattice aspect of Boolean algebras provides links for *ease of access*. (D) The propositional aspect of Boolean algebras, where $\wedge$ stands for 'and', $\vee$ stands for 'or', and $\neg$ stands for 'not' may underlie an intepretation of the representation in logical formulas, and be applied for *ease of inference*.

Likewise, Boolean lattices of connotations *as triggers of reactions* serve purposes of upscaled behavior: (A) The lattice aspect of Boolean algebras provides links for automatic connections and arbitrations between reactions as explained before. (B) $(\imath)$ The binary operation $\wedge$ provides nodes for handling simultaneous reactions, as discussed above. $(\imath\imath)$ The unary operation $\neg$ opens the possibility to model reactions that are conjured in the *absence* of stimuli, when $\varrho(w, \neg\alpha) = t$, and hence $\varrho(w, \alpha) = f$[4]. (Recall that the basic setting models reaction activation only if $\varrho(w, \alpha) = t$.) $(\imath\imath\imath)$ The binary operation $\vee$ opens the possiblity to develop generalized, multi-purpose, reactions that cover a wider range of discriminations. $(\imath\imath\imath\imath)$ $\mathcal{R}(\top)$ would model behaviors that are invariably activated, because $\varrho(w, \top) \equiv t$. Using the same infrastructure, the formalism is thus extended to model permanent activity towards general fixed goals, that is not contingent on specific stimuli. (C) The partial order enables *taxonomic hierarchies* of connotations, hence when one connotation subsumes another connotation, the relevant reactions should subsume one another as well, with inheritance of procedure. (D) The propositional aspect of Boolean algebras, where $\wedge$ stands for 'and', $\vee$ stands for 'or', and $\neg$ stands for 'not' provides infrastructure for rational, 'off line', high-level planning of, and reasoning about, behavior.

(Arzi-Gonczarowski & Lehmann 1998a) studies the mathematical technical aspect of the subcategory of Boolean perceptions, where sets of connotations are Boolean algebras and three-valued p-predicates are embedded adequately, to enable a sensible perception of Boolean combinations of connotations.

[3] Foundationalism, the view that knowledge has a two-tier structure: some of it is foundational (e.g. is justified by sensory or perceptual experiences), while the rest thereof is inferential in that it derives from foundational knowledge, has been widely held from Aristotle, through Descartes, to Russel, Lewis, and most contemporary epistemologists. Likewise, neurologists distinguish between primary and secondary emotions (Damasio 1994).

[4] The formalism is shown to yield a deductive apparatus, that may be algorithmically applied, for the computation of specific values of a three valued Boolean p-predicate, from the values of basic perception.

For rational, context free, reasonings about arbitration between behaviors, one may close the set $\mathcal{Z}$ of behaviors into a lattice. (Whereas Boolean lattices of connotations, as shown above, are about arbitration between behaviors *in the context of specific connotations*).

In the regularity spirit of categorical formalisms, transitions from a basic perception to a perception of a Boolean closure of connotations, with upscaled behavior and representation capabilities, are formalized by p-morphisms. (Arzi-Gonczarowski & Lehmann 1998a) also provides theorems about commutative diagrams that show the methodical equivalence of alternative arrow paths.

Boolean Closures of W-elements

Boolean closures of sets of w-elements, namely environments, are applied to model creative design processes (Arzi-Gonczarowski 1999a). In the proposed formalism, a w-element can be modeled by sets of connotations: the set of connotations that it has, the set of connotations that it does not have, and the set of connotations that are imperceptible/irrelevant for that perception. The computational idea is to construct and to manoeuver subsets of connotations as basis for conceived plans and designs.

Subsets of $\mathcal{I}$ model w-elements in a *conceived environment* of the relevant perception. Obtaining subsets of connotations from Boolean combinations of other subsets of connotations models conception of w-elements on the basis of other w-elements[5]. This formalization opens the possibility for a computational version of the use of examples, similes, and metaphors. One could, for example, specify a combination of similes, stating that some w-element is conceived by a compound resemblance to other w-elements[6].

When the environment is internally conceived, there is no immediate reality to experience and to appreciate. Imaginative design is, indeed, a trying cognitive process that necessitates an 'inner eye'. In the formal context the 'mind's eye' is modeled by a deductive apparatus for the computation of specific values, of the three valued p-predicates, of perceptions with Boolean environments. All that is structurally dual to the construction of perceptions with Boolean sets of connotations.

Upscaled Observation of Lawlike Patterns

As mentioned before, Boolean algebras feature a partial order, enabling the organization of connotations in *taxonomic hierarchies*. A Boolean closure would, of course, place x below $x \vee y$, which is more general. That is a context free Boolean law (there are, of course, others) that always holds. In addition to that, there may be context specific 'law-like'

[5] The use of Boolean operations in planning has already been introduced in STRIPS, and it has been applied in later planning systems as well.

[6] For example, Greek mythological monstrous forms consist of mixtures of attributes from different species: The Centaur horse-man, the Minotaur bull-man, Echidna the snake-woman, Pegasus the horse-bird, Sphinx the woman-lion-bird, Siren the bird-woman, and so on.

patterns, that hold on top of the general logical subsumptions. *All men are mortal* is one famous example. To enhance perceptions with domain specific observational capabilities, ISAAC formalizes two canonical types of transitions to Boolean closures: One is totally context free, while the other provides infrastructure that enables introduction of context specific lawlike patterns into the Boolean structure as well. Lawlike patterns are synonyms and subsumptions among connotations (Arzi-Gonczarowski & Lehmann 1998a), and, dually, congeneric and subjacent w-elements (Arzi-Gonczarowski 1999b; 1999a). Both canonical types of transitions to Boolean closures are formalized by free functors, providing rigorous mathematical descriptions of methodical cognitive transitions to perceptions with inner representations of environments. The mathematical framework provides a detailed comparison between the more general and the more constrained free generations.

The p-morphism tool is upscaled accordingly by relevant set maps that are also Boolean homomorphisms. In addition, p-morphisms can be monotonous with respect to context specific lawlike patterns. Preservation of synonyms and subsumptions among connotations upscale interpretive transitions and representation formations. Preservation of congeneric and subjacent w-elements upscale analogy making and creative design processes. (An independent implementation of a modul that detects lawlike patterns in Boolean algebras actually exists (Boros *et al.* 1996)).

Computational Synthesis, So Far

With a single structuring tool, that consists of Boolean closures of sets, ISAAC has been extended to model behavior integration and control, representation formation, and creative imaginative design processes. A fallout is that representation formation is 'connatural' to design processes. Here one achieves modularity and abstraction by repeatedly applying a single generalized tool to different sets, of different nature, scaling to different high level intelligent capabilities. Only the underlying structure reveals the deep connection between the capabilities that are being modeled, gaining us further insight into intelligent processes.

Theoretical results about the various Boolean closures are captured by commutative diagrams, that show the methodical equivalence of alternative arrow paths. In conventional equations, if the concepts and measurement units of several equations match, then they may be embedded in one another. Like equations, these commutative diagrams are composed into an integrated compound whole because they share vertices and edges in a categorical graph. The integrated commutative diagram provides a high level *blueprint* for the integrated design of intelligent activities (Arzi-Gonczarowski 2000a), perhaps as anticipated by (Magnan & Reyes 1994).

Confines of Boolean Synthesis

Having scaled from $(\imath)$ basic sensory-motor-neural perceptions, to $(\imath\imath)$ perceptions with Boolean structures (i.e. behavior integration and control, high level representation formation, and creative imaginative design capabilities), to $(\imath\imath\imath)$ perceptions enhanced with domain specific observational capabilities, it is natural to ask whether one could do even better with the same Boolean tool of comutational synthesis. The framework provides tools of rigour to systematize intuitions about the confines of minds and intelligence. A *fixed point* theorem answers the question in a precise manner: at this level of abstraction, and with these Boolean tools, the Boolean constructs, enhanced with domain specific observational capabilities, provide the most structured representations and conceived environments that a system could behave upon. The intuitive fallout is that the basic perception, that one uses to generate the relevant Boolean closures, both enables and circumscribes that which could be plausibly represented and conceived. (There are other meaningful bounds: combinatorial bounds, as well as a lax *terminal object* in the category, but the fixed point bound is the strongest (Arzi-Gonczarowski 2000a).)

Scaling in Other Directions

Having exhausted the Boolean tool of computational synthesis, it is still possible that other types of compositions of building blocks and other abstractions could achieve additional high level functionalities.

The form of composition of building blocks that is added now is to let perceptions perceive themselves, as well as other perceptions. Technically, with no need of additional definitions or building blocks, we just let perceptions be w-elements in environments. Intuitively, a modelled perception bends its perceptive binoculars to view others, or itself, as an object that is being perceived. When $\mathcal{P}_1$ perceives $\mathcal{P}_2$, makes discriminations about it, and reacts, then $\mathcal{P}_2$ is a w-element in the environment of $\mathcal{P}_1$, namely $\mathcal{P}_2 \in \mathcal{E}_1$.

This proposal raises theoretically problematic issues that go back to paradoxes which led to an overhaul of the foundations of set theory and modern math. These paradoxes typically originate in self references, or in vicious regress. If $\mathcal{P}_2$ also perceives $\mathcal{P}_1$, the reciprocity introduces circular reference. If, for instance, each one of the behaviors $\mathcal{R}_1, \mathcal{R}_2$ depends on the perception of the other behavior, one gets a vicious circle, that would challenge the *iterative hierarchy* of the construction: Begin with some primitive elements (w-elements, connotations, behaviors), then form all possible perceptions with them, then form all possible perceptions with constituents formed so far, and so on. In set theory, the *axiom of foundation* is normally added to the five original axioms of Zermelo, to warrant an iterative hierarchy.

The main motivation to go ahead and formalize perceptions of perceptions anyhow, where, for example $\mathcal{E}$ is allowed to be a 'non classical' set (Aczel 1987), is that *these are precisely the theoretical difficulties, that are inherent in the construction, that model difficulties of self perception and the perception of others.* Vicious circles do happen in self reflection and in social situations, and they need to be modeled. An agent could recur into infinite regress, perceiving itself as it perceives itself, and so on, requiring more and more resources and eventually derailing the system. (Sloman 2000) classifies reflective emotions together with other perturbant states that involve partly losing control of thought processes. He also remarks that: *'Self-monitoring, self-evaluation, and self-control are all fallible. No System can have full access to all its internal states and processes, on* *pain of infinite regress'*. That is a 'no free lunch' price: If a theoretical model of self reflective and social intelligence had consisted of straight line computations that always converge, then that would have provided a major reason for serious worries concerning the *validity* of that model.

It should be noted that not all self references produce theoretical paradoxes, just as not all perceptions of perceptions involve vicious regress. Some are benign and bottom out neatly. The philosophical and mathematical difficulties lie in forming conditions that exclude the pathological cases *only*.

(Arzi-Gonczarowski 2001a; 2001b) discuss how to capture, within ISAAC, self reflection, with its unavoidable entanglements, and social cognition, with its unavoidable irksome deadlocks, by application of a *nonwellfounded* mechanism of *higher order perception*.

summary

ISAAC models perceptual, cognitive, and affective, intelligent behavior by computational synthesis of a relatively small number of mathematical building blocks. Like a reduced instruction set for a RISC computer, the basic building blocks, supported by known mathematical Boolean and categorical constructions, are reused in various ways to scale to high complexities and to model high level functions of intelligence, that is general purpose and domain independent. ISAAC's approach predicts tidily structured implementations, featuring modularity, regularity, abstraction, and awareness of hierarchy. The advantage of these design principles is more then mere technical elegance, suggesting a theoretical standard and neat implementations — it also means that the premises do capture basic issues that are shared by intelligent processes.

References

Aczel, P. 1987. *Lectures in Nonwellfounded Sets*. Number 9 in CSLI Lecture notes. CSLI.

Allen, J. 1998. Ai growing up – the changes and opportunities. *AI Magazine* 19(4):13–23.

Arzi-Gonczarowski, Z., and Lehmann, D. 1998a. From environments to representations–a mathematical theory of artificial perceptions. *Artificial Intelligence* 102(2):187–247.

Arzi-Gonczarowski, Z., and Lehmann, D. 1998b. Introducing the mathematical category of artificial perceptions. *Annals of Mathematics and Artificial Intelligence* 23(3,4):267–298.

Arzi-Gonczarowski, Z. 1998. Wisely non rational – a categorical view of emotional cognitive artificial perceptions. In Cañamero, D., ed., *1998 AAAI Fall Symposium: Emotional and Intelligent*, 7–12.

Arzi-Gonczarowski, Z. 1999a. Categorical tools for perceptive design: Formalizing the artificial inner eye. In Gero, J., and Maher, M., eds., *Computational Models of Creative Design IV*. University of Sydney, Australia: KCDCC. 321–354.

Arzi-Gonczarowski, Z. 1999b. Perceive this as that - analogies, artificial perception, and category theory. *Annals of Mathematics and Artificial Intelligence* 26(1-4):215–252.

Arzi-Gonczarowski, Z. 2000a. A blueprint for a mind by a categorical commutative diagram. In *Proceedings of the AISB'00 Symposium on How to Design a Functioning Mind*, 10–18.

Arzi-Gonczarowski, Z. 2000b. A categorization of autonomous action tendencies: The mathematics of emotions. In *EMCSR 2000*, volume 2, 683–688.

Arzi-Gonczarowski, Z. 2001a. Perceptions that perceive themselves – a mathematical schema. *International Journal of Computing Anticipatory Systems* 8:33–51. Invited paper.

Arzi-Gonczarowski, Z. 2001b. Self, empath, manipulativity: Mathematical connections between higher order perception, emotions, and social cognition. In Cañamero, D., ed., *2001 AAAI Fall Symposium: Emotional and Intelligent II*, 9–14.

Arzi-Gonczarowski, Z. 2002. Ai emotions: Will one know them when one sees them? In *EMCSR 2002*, volume 2, 739–744.

Barr, M., and Wells, C. 1995. *Category Theory for Computing Science*. Prentice Hall.

Boole, G. 1854. *An Investigation of The Laws of Thought*. Macmillan. Reprinted by Dover Publications, N.Y., in 1958.

Boros, E.; Hammer, P.; Ibaraki, T.; Kogan, A.; Mayoraz, E.; and Muchnik, I. 1996. An implementation of logical analysis of data. Rutcor Research Report RRR 22-96, Rutgers University, New Brunswick, N.J.

Damasio, A. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York, NY: Gosset/Putnam Press.

Frijda, N. H. 1986. *The Emotions*. Cambridge: Cambridge University Press.

Gould, S., and Vrba, E. 1982. Exaptation: A missing term in evolutionary theory. *Paleobiology* 8:4–15.

Herrlich, H., and Strecker, G. 1973. *Category Theory*. Allyn and Bacon.

Lawvere, F., and Schanuel, S. 1997. *Conceptual Mathematics*. Cambridge University Press.

MacLane, S. 1972. *Categories for the Working Mathematician*. Springer-Verlag.

Magnan, F., and Reyes, G. 1994. Category theory as a conceptual tool in the study of cognition. In Macnamara, J., and Reyes, G., eds., *The Logical Foundations of Cognition*. Oxford University Press. 57–90.

Sikorski, R. 1964. *Boolean Algebras*. Springer Verlag.

Sloman, A. 2000. Architectural requirements for human-like agents, both natural and artificial (what sorts of machines can love?). In Dautenhahn, K., ed., *Human Cognition and Social Agent Technology*. John Benjamins Publishing. 163–195.