Modular Assembly of Intelligence by Mathematical Abstractions

Zippora Arzi-Gonczarowski
Typographics, Ltd.
46 Hehalutz Street
Jerusalem 96222, Israel
zippie@actcom.co.il
www.actcom.co.il/typographics/zippie

Abstract

A mathematical abstraction is applied to model intelligence from the bottom up. Starting from an abstraction of a basic embodied match between sensations and reactions, the modelling process then integrates a relatively small number of modular building blocks to scale along a continuum from low- to general high-level intelligence. The distal goals are manageable implementations, a theoretical standard, a unified ontology and language of discourse about embodied intelligence, and insights into basic issues that are shared by processes and forms of intelligence.

Introduction

Building high-level intelligent systems would indeed require a potpourri of technologies. The question whether intelligence is merely a disparate collection of things that are to be packed together, or maybe they can be modelled with shared concepts and moduls, is not only a concern regarding their practical integration in a working system. This question is also scientifically interesting: Resolving collections of phenomena into theoretical units of abstraction, thus modelling them with shared concepts, has always been a hallmark of scientific visions. It allows an analytic, deep understanding of the subject matter. (The ability to compare AI architectures and systems would be just one example fallout of that.)

This is a proposal for a unified view of intelligence, outlined for the community of the workshop on modular construction of human-like intelligence. The underlying idea is to find a more suitable level of abstraction, higher than the level that one is perhaps used to: Step back and look for structure in the muddle. Perhaps, for instance, one could describe some ad hoc solutions by decomposition into sub modules, abstract module functionality, and find a unifying framework there. The tool is mathematics, the lingua franca of science, which is the human endeavor that was developed to deal with abstraction (Devlin 2003).

Risks in climbing up the ladder of abstraction are either being lost in clouds of generalizations, or, when clinging on to a meaningful distinction, stumbling down and being lost in a mass of detail. The challenge is a balance between abstraction that is not detached, and grounding that is not over deterministic. Mathematical categorization has been developed for such purposes within mathematics itself.

Readers are invited to a careful choice and formalization of modular building blocks, their rigorous composition into useful modules, abstraction of module functionality, then further reusage. All that while giving one's mind to regularity, symmetry, hierarchy, and further principles that minimize information content[1] and result in an integrated, manageable, vision of intelligence, hopefully getting us closer to a scientific vision of a Big Picture.

Methodology and Background

The essence of mathematical modelling has always been to start from basic, low-level, building blocks which are intuitively convincing and obvious. Then, following a long consecution of simple steps, each one intuitively convincing by itself, to obtain arbitrary high-level constructs. One typical paradigm is the system of natural numbers: The five postulates of Peano capture the pre-theoretical essence of the natural numbers as counters. Orderly extensions of the natural numbers provide the integers, then the rational numbers, then the real numbers. Hilbert coined the term *The Genetic Method* for the method which is suggestive in the present context. 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, from the simple axioms, highly nontrivial knowledge. Another paradigm is Euclidean geometry, modelling even the most intricate structures of our physical space starting from points and lines as modular building blocks.

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). Intelligent systems could be modeled mathematically by starting from primitive modular building blocks that capture an abstraction of that, and orderly structured extensions could then be introduced to model higher level funcionalities, applying deduction to

[1] See (Lipson, Antonsson, & Koza 2003) for principles of computational synthesis

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.

ISAAC, an Integrated Schema for Affective Artificial Cognition, is a formalism that follows the above guidelines. The formalism boots agents' 'minds' from a formalization of mutually driven perceptions and reactions, abstracting an essence of the natural evolutionary context, thus also capturing the essence of embodiment (Anderson 2003a; Chrisley 2003; Anderson 2003b). Modular compositions then scale up all the way to rational, abstract thinking. Complex high level intelligence is thus modeled using a rigorous mathematical framework and a circumbscribed number of reusable and composable modular constructs, yielding a continuum from low- to high- level intelligence, and a unified schema that models an integrated view of a wide range of intelligent processes.

The scope of this outline of ISAAC for the workshop on modular construction of human-like intelligence, limits us to an intuitive, hand waving, synopsis of the mathematical formalism. Readers interested in theorem proofs, technical details, and grounding examples, are encouraged to refer to the published papers that are available at the author's web page (Arzi-Gonczarowski 2005).

ISAAC's Basic Modular 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 modular 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 may not be 'aware of its concerns'). Behaviors are typically conjured up as responses to stimuli in the environment, hence agents are provided with a sensing apparatus, and the modular building blocks should account for that. From these bare basics the formalism is going to proceed, heel and toe, to higher-level, nontrivial, complex manifestations of human like intelligence.

Natural evolution selected sensory motor neural apparatuses that coupled embodiments of organisms with their ecological niches, yielding behavior that could be 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 by the system.

It has long been accepted that forms of natural intelligence, including sublime human ones, are results of (combinations of) improvements and upgrades applied, by natural selection, on top of basic reactive intelligence. But 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 modular building blocks that stand for basic discriminating sensations and basic reactions that go with them. In the service of domain independent modularity and mathematical abstraction, the specific grounded instantiation of these modular building blocks is left undetermined: an agent's perceptive-reactive state ('*p-r state*') $\mathcal{P}$ makes a discrimination α about some world chunk w, and that conjures up a reaction r (a formal definition follows). Substitution instances of the schema will provide these deliberately meaningless symbols a concrete substantiality. Rather than dodging the embodied grounding issue, it is entered into the formalism.

Based on these modular building blocks, and following a series of rigorous compositional steps, the proposed theory obtains high-level cognitive, behavioral, and affective modular constructs and functions of intelligence. Abstraction of modul functionality and reusage of moduls occur naturally along the way, providing us with positive feedback that the proposed premises may be useful and adequate for an integrated modular view of general high-level intelligence.

A Basic Modul

All the publications about the proposed formalism share the following premises, with the longer ones featuring extensive discussions of the methodical considerations behind these premises, as well as elaborated examples[2].

Definition: A *Perceptive-Reactive State* ('*p-r state*') 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 $\varrho : \mathcal{E} \times \mathcal{I} \longrightarrow \{t, f, u\}$.
- $\mathcal{R}$ is a function: $\mathcal{R} : \mathcal{I} \longrightarrow \mathcal{Z}$

[2]In the course of 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 the formal basis for the proposed theory. Author's web page (Arzi-Gonczarowski 2005).

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. Specific instantiations of w-elements in specific states would come from the open-ended diversity of formalized ways to capture indexicals in a system's context: spatial or temporal coordinates, 'the thing that sensor x bears upon', and so on. Later, the concept of perceived environments will be abstracted, to be reused also for conceived environments and w-elements that are just imagined, or recalled, and further on to self referral, thus reusing and scaling the same modular definition to model higher capabilities of intelligence.

The elements of $\mathcal{I}$ model sensed discriminations that are afforded by the perceiving agent[3]. Each discrimination models a 'socket' where a sensed stimulus 'plugs in' (by the perception predicate ϱ as described below). For example, if the environment of a p-r state features a w-element ALARM_BELL that happens to be sounding, perception could 'plug it' into the discrimination *alarm_sound* (and possibly also to other discriminations, such as *loud_sound* or *interrupted_sound*). (At this point it is appropriate to recall Gibson's *affordances* (Gibson 1977): the resources that the environment offers an animal, and it needs to possess the capabilities to perceive them and to use them.) When a reaction is 'wired' to a discrimination, it would be conjured up (that is modeled by $\mathcal{R}$, described below). Later, the role of discriminations will be exapted to labeled representations that stand for the relevant stimuli. That way the system will be able to access and summon discriminations by internal 'thought' processes. (With a capability to interrupt automatic connections between discriminations and reactions.) Again, the same modul will be reused and scaled to model representational capabilities of higher intelligence.

The 3-valued *Perception Predicate* (*p-predicate*), ϱ, models perception of discriminations in w-elements: $\varrho(w, \alpha) = t$ stands for definitely 'yes', namely 'w has the discrimination α'. $\varrho(w, \alpha) = f$ stands for definitely 'no', namely 'w does not have the discrimination 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 $\varrho(w, alarm\_sound) = t$, If the agent perceives that the bell is quiet then one gets $\varrho(w, alarm\_sound) = f$. Lastly, $\varrho(w, alarm\_sound) = u$, 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 p-r state can activate, modeling a set of behaviors, actions, and reactions, for that p-r state. For basic reactive agents, it would typically consist of that agent's set of physical reactions (`feed`, `fight`, or `flight` are common examples). An element `null` in $\mathcal{Z}$ may stand for the empty call, namely no response, or indifference. Later, the concept of actions will be extended to introvert mental activity, to innately motivated activity that is not necessarily triggered by outside stimuli, and to activities that control other activities. Among others, that provides the premises for a modular systems approach to modeling emotions, affect, and related phenomena. The same modul 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: $\mathcal{R}(alarm\_sound) =$ `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 modul for the modeling of 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.

Specific instantiations of the five coordinates from the definition: $\mathcal{E}$, $\mathcal{I}$, ϱ, $\mathcal{Z}$, and $\mathcal{R}$, provide a specific p-r state. The formalized modul $\mathcal{P}$ hence stands for any basic embodied reactive precognition. It is high-level in the sense that it is presumed to layer on top, and be grounded by, a sensory motor neural apparatus. It is low level (in another sense) if these coordinates never change. In that simplistic case our story ends here: These p-r states could then be easily programmed using a loop that checks the sensors and reacts accordingly, practically conflating ϱ with $\mathcal{R}$. But ISAAC proceeds beyond that, as explained below.

Basic Modular Dynamics

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 one 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 p-r state into another. Flows of conversion between p-r states are formalized by *perception morphisms (p-morphisms, arrows)*:

Definition:
Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two p-r states:

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

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

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

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).

[3] in the earlier papers they were named *connotations*.

The set maps are formallizations of transitions between perceived environment, or changes in discriminations, or changes in behavior, respectively. Mathematical set maps afford 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 discrimination, 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, discriminations, behaviors) may be merged, modeling amalgamations, generalizations, combinations, etc. Set maps also enable conversion of one constituent into another. Replaced discriminations typically indicate translations or interpretations of discriminations, replaced w-elements typically indicate similarity of objects in some respect, (such as in analogies), and replaced behaviors typically indicate adaptation of reactions.

These p-morphisms are 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 discriminations, and, on the other hand, transitions between discriminations need to be grounded by commensurate experience, and factorization of arrows may keep 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 p-r states 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. 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. Results that have not been anticipated at the outset provide supporting arguments that the proposal is apparently on a promising track.

It is sometimes helpful to consider a category, and $\mathcal{P}rc$ in particular, as a graph. In that underlying structure, p-r states are vertices and p-morphism arrows are edges. Formalized intelligence processes perform by construction and navigation of relevant portions of that graph. Transitions between p-r states 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[4].

Modular Integration, 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 p-r states. 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[5]. A target p-r state of some p-morphism could even, for example, exist prior to the domain p-r state 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' p-r states provides basis for modeling paths of communication. In that case, both the domain and the target p-r state 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.)

A modular regularity of intelligent processes in ISAAC is that they consist of (structured compositions of) p-morphisms: Interpretive transitions, representation formations, analogy making, creative design, intra-agent communications, joint p-r states, and more.

Having introduced arrows between p-r states, one can immediately integrate this modular building block into the framework. In the basic definition of p-r states, the set $\mathcal{Z}$ models a collection of behaviors for that state. 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

[4](Barr & Wells 1995, p.83) entitle commutative diagrams as *the categorist's way of expressing equations*.

[5]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.

response to some stimulus.

The implications of this last composition of modular building blocks could be far reaching, and it has the potential of scaling the system to surprising complexities. Since p-r states, as defined, determine reactions, a transition $h : \mathcal{P}_1 \rightarrow \mathcal{P}_2$ 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' p-r states, with minimal constituents as bootstrap. (All arrows are bound on their left by the *empty p-r state*, the initial object of the category, modeling a theoretical *Tabula rasa*.) If a system is 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. (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 a different p-r state.)

P-morphisms are themselves made of regular modules: There is a structural similarity between the $\mathcal{I}$ mapping ($h : \mathcal{I}_1 \rightarrow \mathcal{I}_2$) of a p-morphism as the *interpretive* component of the transition, and the $\mathcal{E}$ mapping ($h : \mathcal{E}_1 \rightarrow \mathcal{E}_2$) 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 modular 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 modular building blocks, and maps between them. Computationally abstracted, both are set maps. A salient property of the p-morphism modular construct is the symmetry between w-elements and discriminations as variables of the p-predicate. From a purely technical, context free, point of view, their roles are interchangeable. This *duality* has modular consequences, both theoretical and computational. For example, any formal construction or theorem that is established for discriminations (w-elements) can automatically be applied to w-elements (discriminations), mutatis mutandis. This suggests theoretical insights into a 'connaturality' of processes and capabilities. This entails architectural and applicational modularities.

Higher Integrated Moduls

Except for trivial contexts, any p-r state 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 probably with 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 discriminations $\mathcal{I}$, or w-elements $\mathcal{E}$) under Boolean connectives, provides infrastructure for achieving a wide range of high-level intelligent functionalities. Based on results of the well developed theories of Boolean algebras and of categories, a methodical closure of constituents into Boolean lattices has been formulated. ISAAC applies this abstract procedure to the various constituents of p-r states. (The p-morphism tool is upscaled accordingly by letting relevant set maps be Boolean homomorphisms.)

Lattices of Discriminations and Related Reactions

In the basic definition, whenever perception detects in its environment a w-element with a certain discrimination (as defined by ϱ), perception 'plugs it' into the 'socket' associated with that discrimination, triggering reactions that are 'wired' to that 'socket' (as defined by $\mathcal{R}$). At that basic level, if more than one discrimination 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 that all these discriminations were interconnected and arranged in a lattice, where every combination of discriminations corresponds to a junction node in the lattice. Since the basic reactions are innate, they are still 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, however, easy to see that most solutions to activity conflicts are likely to 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 resources are being invested in con-

taining that process. Similarities between that and human emotions have been shown in the context of ISAAC. 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 exapted: while for lower level systems it models reactions that are invariably performed, in higher level ones it could also model action tendencies that are not consummated. The result is upscaling the reactive capability to also include a certain sense of emoting, as well as a certain sense of self control, in higher level intelligences.

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. As mentioned earlier, resources may be required for that kind of control, and the formalization of that within ISAAC naturally integrates the modeling of emotions and affect into the framework.

A behavioral pressure for the introduction of lattices of discriminations has just been described. This is now followed by showing how these lattices can be modularily 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 modular building blocks that already discriminate: When discriminations are referrable, then 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 discriminations, could serve representational purposes and related procedural objectives: (A) They feature a partial order. This may enable the organization of discriminations 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 [6]. (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 discriminations *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 $\varrho(w, \alpha) = f$ [7]. (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. $(\imath\imath\imath)$ $\mathcal{R}(\top)$ 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 discriminations, hence when one discrimination subsumes another discrimination, 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.

The mathematical technical aspect of the above is captured in the subcategory of Boolean p-r states, where sets of discriminations are Boolean algebras and three-valued p-predicates are embedded adequately, to enable a sensible perception of Boolean combinations of discriminations. In the modular regularity spirit of categorical formalisms, transitions from a basic p-r state to a p-r state of a Boolean closure of discriminations, with upscaled behavior and representation capabilities, are formalized by p-morphisms.

Boolean Closures of W-elements

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

Subsets of $\mathcal{I}$ model w-elements in a *conceived environment* of the relevant p-r state. Obtaining subsets of discriminations from Boolean combinations of other subsets of discriminations models conception of w-elements on the basis of other w-elements. This formalization opens the possibility for a computational version of the use of examples, similes, and metaphors. One could, for example, specify a

[6] 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. Analogously, neurologists distinguish between primary and secondary emotions (Damasio 1994).

[7] The formalism is also 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.

combination of similes, stating that some w-element is conceived by a compound resemblance to other w-elements[8].

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 p-r states with Boolean environments. All that is structurally dual to the construction of p-r states with Boolean sets of discriminations.

Upscaled Observation of Lawlike Patterns

As mentioned before, Boolean algebras feature a partial order, enabling the organization of discriminations 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' patterns, that hold on top of the general logical subsumptions. *All men are mortal* is one famous example. To enhance p-r states 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 discriminations, and, dually, congeneric and subjacent w-elements. Both canonical types of transitions to Boolean closures are formalized by free functors, providing rigorous mathematical descriptions of methodical cognitive transitions to p-r states 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 discriminations 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)).

Modular Integration, 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 theoretical 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 modular 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 modular design of intelligent activities, perhaps as anticipated by (Magnan & Reyes 1994).

Confines of Boolean Integration

Having scaled from $(\imath)$ basic sensory-motor-neural p-r states, to $(\imath\imath)$ p-r states with Boolean structures (namely: behavior integration and control, high level representation formation, and creative imaginative design capabilities), to $(\imath\imath\imath)$ p-r states enhanced with domain specific observational capabilities, it is natural to ask whether one could do even better with the same Boolean tool of modular integration. The mathematical framework provides tools of rigour to systematize intuitions about the confines of that: 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 p-r state, 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.)

Integration in Other Directions

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

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

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 $\mathcal{P}_j$

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

also perceives $\mathcal{P}_i$ (e.g. if $i = j$), the reciprocity introduces circular reference. If, for instance, each one of the behaviors $\mathcal{R}_i, \mathcal{R}_j$ depends on the p-r state 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, discriminations, behaviors), then form all possible p-r states with them, then form all possible p-r states with constituents formed so far, and so on. The *axiom of foundation* is normally added to the five original axioms of set theory, to warrant an iterative hierarchy.

Still, it is possible to go ahead and formalize p-r states of p-r states anyhow, where $\mathcal{E}$, for example, is allowed to be a 'non classical' set (Aczel 1987). One motivation being 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*.)

ISAAC's Grounding and Testing

In the tradition of the scientific paradigm, the hypothesis is that existing intelligent systems, natural and artificial, can be dissected, analyzed, and explained with the proposed principles and fundamental recurrencs, and the prediction is that novel and manageable implementations can be neatly synthesized, designed, and constructed using them. (The published papers that are available at the author's web page (Arzi-Gonczarowski 2005) provide quite a few elaborated grounding examples.)

Summary

ISAAC offers to model upscaled intelligences mathematically, by modular integration of a relatively small number of abstracted modular building blocks. Like a reduced instruction set for a RISC computer, the basic modular building blocks, supported by known mathematical Boolean and categorical compositions, can be regularily integrated and reused, in various systematic ways, to scale along a continuum from low- to high-level functions of intelligence, that is general purpose and domain independent. ISAAC's approach offers insights into principles and fundamental recurrences that are shared by processes and forms of intelligence. It hypothesizes a theoretical standard and a unified ontology and language of discourse to dissect and to construct embodied intelligences, to analyze and to synthesize, to explain and to compare them.

References

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

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

Anderson, M. L. 2003a. Embodied cognition: A field guide. *Artificial Intelligence* 149(1):91–130.

Anderson, M. L. 2003b. Representations, symbols, and embodiment. *Artificial Intelligence* 149(1):151–156.

Arzi-Gonczarowski, Z. 2005. Home page: Papers. http://www.actcom.co.il/typographics/zippie.

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. L.; 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.

Chrisley, R. 2003. Embodied artificial intelligence. *Artificial Intelligence* 149(1):131–150.

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

Devlin, K., ed. 2003. *Why* CS *Students Need Math (Special Section)*. Communications of the ACM 46(9). 37–56.

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

Gibson, J. J. 1977. The theory of affordances. In Shaw, R., and Bransford, J., eds., *Perceiving, Acting, and Knowing*. New-York: Wiley. 67–82.

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

Lipson, H.; Antonsson, E. K.; and Koza, J. R., eds. 2003. *Computational Synthesis: From Basic Building Blocks to High Level Functionality*. AAAI Press Technical Report SS-03-02.

Magnan, F., and Reyes, G. E. 1994. Category theory as a conceptual tool in the study of cognition. In Macnamara, J., and Reyes, G. E., 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.