Perceptions that Perceive Themselves

Zippora Arzi–Gonczarowski
Typographics Ltd, Jerusalem 96222, Israel

2001 draft

Abstract

It is shown how some aspects of consciousness can be structured mathematically. This is based on the category of artificial perceptions, that has been conceived and proposed as an infrastructure for a theory of AI processes, and as a high level AI architecture. This unified theoretical standard has already been shown to enable a rigorous interweaving and integration of various intelligent capabilities, and in this paper self reflective capabilities are also integrated into the schema.

1 Introduction

A general concept of intelligent mental activities, namely the mind, typically involves the complex organized whole of the individual's adaptive activity, including perceiving, remembering, considering, conceiving, reasoning, evaluating, deciding, and the like. These activities are, in philosophical terminology, intentional: they are *about* the things that are being perceived, considered, etc. These are, first and foremost, entities in the environment of the agent under consideration. Common examples are physical objects, other agents, physical and social events that involve those, etc. When one conceives of a mentally capable AI artefact, these would often be the issues considered.

We are still far from achieving the long term goal of designing a functioning artificial 'mind' that integrates the above (partial) list of activities. However, various research in AI and related disciplines has been able to come up with proposed architectures and with analyses of requirements for divers aspects of such architectures (Sloman, 2000 [26]). Let us use these preliminary steps as basis for a figment of the imagination, and assume for a moment that all this has already been modelled. This paper is about what would happen if a modelled 'mind' were able to bend its perceptive binoculars to view *itself* as the object that is being perceived, evaluated, etc. More specifically, instead of relating to X which is an external entity, the 'mind' is allowed to substitute its own workings for X. A few questions that could be asked:

1. Under what conditions is this theoretically valid.
2. Is this an interesting topic of study: would it extend or improve agent's capabilities?
3. What natural mind capabilities does this capture, if at all.
4. To what extent could the capability to relate to its own self as an object, be phenomenally grounded in an AI artefact.
5. What would be the limitations or undesired effects of that.

To ground the discussion, we take off from a mathematical model of mental processes that has been proposed and discussed in a series of publications. To make the discussion somewhat self contained, the first part of the paper provides an outline of that model, and briefly summarizes the published results. It will be argued that the mathematical categorical schema captures an essence of mental processes, yet it avoids over determinism and is hence general enough for its purpose. (Mathematical category theory typically provides formal tools to capture a structural essence without being over deterministic.) The second part of the paper constitutes its novel import, integrating into the high level categorical schema the self reference that has been proposed above, offering tentative answers to the questions asked. As is sometimes the case with new research directions, answers are also provided to questions that have not been initially raised. The capability to perceive one's own mental apparatus, here and now, requires conceptualizations that can also be applied to other minds and states of mind, and to processes other than the current ones. Some fallout capabilities will be discussed as well.

1.1 A word about consciousness

Already in 1690, John Locke defined consciousness as '*the perception of what passes in a man's own mind*'. Hence, if one manages to provide an artificial artefact with the possibility to perceive, cognize, and emote about its own perceptual cognitive and affective processes, this may perhaps approximate some aspects of consciousness.

Consciousness has been an intriguing topic in philosophy for centuries, but only later in the 20th century it started to draw attention also from psychologists, neuroscientists, evolution biologists, and AI researchers. In the last decade it has finally become a legitimate topic of scientific inquiry, with an ever increasing number of publications. It is beyond the scope of this paper to do justice to the vast body of thought and research about consciousness. A methodical survey of a variety of approaches to consciousness is provided by Seager (1999 [23]). A partial list of influential modern theorists of consciousness should include, among many others, Ned Block, David Chalmers, Patricia and Paul Churchland, Robert Cummins, Antonio Damasio, Donald Davidson, Martin Davies, Daniel Dennett, Fred Dretske, Gerald Edelman, Owen Flanagan, Jerry Fodor, William Lycan, Colin McGinn, Thomas Nagel, John Searle, and David Rosenthal.

The purpose of this study is neither to say new things, that have not been said before, about consciousness, nor is it to argue with, or arbitrate between, different theories of consciousness. The purpose is to show how some aspects of consciousness can be rigorously described mathematically. Besides an eventual contribution of mathematical modelling to scientific understanding, this is a proposal for a high level, computational, AI architecture for agents with perceptual affective and cognitive capabilities, that further deploys the same basic capabilities to support aspects of consciousness.

Consciousness is not a specific phenomenon. It is sometimes said to be a *cluster concept.* It encompasses a broad spectrum of mental capabilities, and there is no general agreement about its definition and boundaries. Various researchers have come up with valuable distinctions: phenomenal consciousness, self consciousness, access consciousness, monitoring consciousness, higher-order

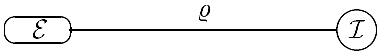

Figure 1: A perception schema

thought, and more. Some of those will be mentioned in context. The ontology that emerges from the schema that is described in this paper often seems to agree with the discriminations of Damasio (1999 [15]), and this will also be pointed out at the relevant places.

2 ISAAC - An Integrated Schema of Affective Artificial Cognition

The schema is based on a mathematical category of *perceptions*, and *perception morphisms* formalize cognitive processes as transitions between perceptions. A categorical approach deals with the nature and the structure of processes and not with particular instantiations. Specific sensory-motor-neural appartuses, coupled with specific environments, should provide substitution instances of the schema. Commutative diagrams offer 'circuit blueprints'[1] for various perceptual cognitive and affective processes.

This section summarizes material from references [3] through [9], each formalizing a different kind of perceptual cognitive process. These references provide examples and discuss the pre-theory, the theory, and the results, at length, while the synopsis below is rather skeletal.

2.1 Basic Constructs

The intentional nature of mind processes, their *aboutness*, is behind the decision to bootstrap the schema from perception. Allen (1998 [2]), for instance, says: *'a prerequisite for something to be intelligent is that it has some way of sensing the environment and then selecting and performing actions.'* High level perception is then based on a classification of environmental chunks.

A *Perception* is defined as a 3-tuple $\mathcal{P} = \langle \mathcal{E}, \mathcal{I}, \varrho \rangle$ where $\mathcal{E}$ and $\mathcal{I}$ are finite, disjoint sets, and ϱ is a 3-valued predicate $\varrho : \mathcal{E} \times \mathcal{I} \to \{t, f, u\}$. The set $\mathcal{E}$ represents the perceived environment, *world elements* (*w-elements*) that could perhaps be discerned by a perceiving artefact. The set $\mathcal{I}$ stands for the internal labels of regularities in w-elements, *connotations* that have a subjective existence that is dependent on the perceiving artefact. The 3-valued *Perception Predicate* (*p-predicate*) ϱ relates w-elements and connotations. Actual sets $\mathcal{E}$ and $\mathcal{I}$, and the values of ϱ, once given, provide a substitution instance, capturing the intuition that perceptions and sensations are innate to artefacts, and develop relative to their environments. The $\mathcal{P}$'s stand for embodied perceptual states. They are high-level, happening at and above the level of recognition of cohesive wholes, where conscious cognizance begins to play a role. Perceptions vary across agents, situations, etc. The diagrammatic description of a perception will be based on figure 1: An oval designates a set of w-elements, a circle

---

[1] The terminology is borrowed from Magnan and Reyes (1994 [21]), who suggest that categorical constructs provide *blueprints* for the design of cognitive activities.

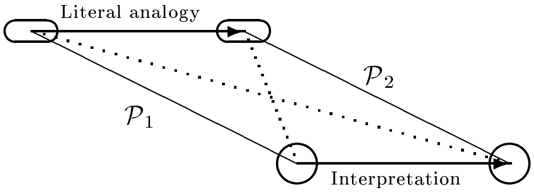

Figure 2: A transition between two perceptions

designates a set of connotations, and the connecting thin line represents some predicative connection ϱ between the two.

Behaviour at the level of this definition was introduced in (Arzi-Gonczarowski, 1998 [5]). Typically, it has to do with survivability. It consists of emotive reactions that are conjured by perception, providing for agents that can not only passively perceive, but also respond and interact with the environment. In (object oriented) programming terminology: for every connotation $\alpha \in \mathcal{I}$ and w-element $w \in \mathcal{E}$, the combination of α, w, and $\varrho(w, \alpha)$ could send a message to an *object*. *Methods that are activated* by these messages are the reactions that are associated with perception, and they are part of the definition of $\mathcal{P}$. This formalism is most typically required to capture wired reactive physical behaviour of survival (like self defence, food consumption, reproduction). All agree that this is the evolutionary origin of the emotions of biological agents. However, once the 'circuitry' for reactions is provided, it may serve action tendencies beyond fundamental urges. Researchers of biological evolution use the term *exaptations* for novelties that arise as features acquired in one context before being co-opted in a different one (Gould and Vrba, 1982 [19]). A wiring could 'exapt' by being connected to *anything that the artefact is capable of doing*. For example, in agents that feature memory or rational capabilities, reactions could be wired to update or retrieval of *data members*, delegation of tasks to higher-level rational procedures, etc. 'Think!' or 'Remember!' could come to be wired, not just 'Fight!' or 'Flight!'.

The flow between perceptions is formalized by p-morphisms (*perception morphisms*): If $\mathcal{P}_1 = \langle \mathcal{E}_1, \mathcal{I}_1, \varrho_1 \rangle$ and $\mathcal{P}_2 = \langle \mathcal{E}_1, \mathcal{I}_2, \varrho_2 \rangle$ are perceptions, then a p-morphism $h : \mathcal{P}_1 \to \mathcal{P}_2$ defines the set mappings: $h : \mathcal{E}_1 \to \mathcal{E}_2$, $h : \mathcal{I}_1 \to \mathcal{I}_2$, and *No–Blur* is the structure preservation condition: for all w in $\mathcal{E}$, α in $\mathcal{I}$, whenever $\varrho_1(w, \alpha) \neq u$ then $\varrho_2(h(w), h(\alpha)) = \varrho_1(w, \alpha)$. The diagrammatic description of p-morphism transitions consists of thick lines between sets of w-elements and between sets of connotations as in figure 2. Every such transition can be factorized into an *interpretation*, which consists of the mapping of connotations, and a *literal analogy*, which consists of the mapping of environments. They can be composed in any order. That is why they are shown as parallels in the figure. Whether the interpretation (or the literal analogy) is the first or the second factor effects the *metaphorical perception* that is generated in between. The dotted diagonals in figure 2 designate the metaphorical perceptions that blend connotations from one perception with w-elements from another.

A transition from $\mathcal{P}_1$ to $\mathcal{P}_2$ may also involve a change in some, or all, reactions, featuring a change of mood or attitude. Moreover, since a reaction could be wired to anything that the artefact is capable of doing, this may also be the *activation of a p-morphism*: perceiving something in the environment can be wired to an internal transition. An example transition may occur when

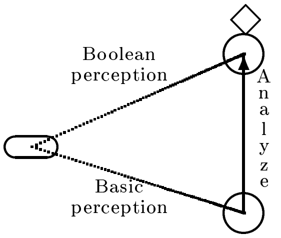

Figure 3: Boolean Representation Generation

an agent perceives how the environment responds to one of its behaviours, and it is impelled to undergo an internal transition to a modified perceptual state that features that behaviour reinforced (or mellowed) according to the perceived response.

Composition and the identity are defined by those of set mappings. A theorem shows that perceptions with p-morphisms make a mathematical category, designated $\mathcal{P}rc$, providing a well developed infrastructure for a mathematical theory.

2.2 Higher Level Constructs

Boolean constructs were applied to further develop a theory. Figuratively, the 'plane' that is shown in figure 2 is going to serve as a 'base' for a commutative diagram that looks like a 'box'. The construction of two 'supporting walls' will be summarized now.

Analytic organizations of grounded representations were formalized by *Boolean generations*, that close sets of connotations under Boolean operations, transforming the $\mathcal{I}$'s into Boolean algebras. (With an adequate restriction of the 3 valued p-predicate for these perceptions.) P-morphism transitions are then based on Boolean homomorphisms between connotations, capturing acute, structure aligning, interpretations. Category theoretical natural transformations systematized the transitions into perceptions that feature the Boolean property. A Boolean combination of connotations is interpretable as a logical formula, so that higher-level reasoning moduls could take this representation as input. It follows that the concreteness of the basic perceptual apparatus is married with the powers of abstraction and the rational capabilities of the higher-level apparatus. The transition is schematized in figure 3, where the Boolean set of connotations is topped with a diamond. The arrow marked *analyze* designates the natural transformation[2]. Everything here pertains to a single environment (the oval). The discussion will be extended later to more environments.

The import of the Boolean construct to behaviour are more advanced levels of autonomy and control. Boolean combinations of perceptual constituents are themselves perceptual constituents, that can be wired to reactions, as in simple perceptions. A possibility follows to couple arbitrarily complex combinations of perceptual constituents with whatever actions, overt as well as introvert, that the agent can effect. The most obvious application is an autonomous regulatory control of emotional conflicts: A complex combination of perceptual constituents may eventually be wired to a complex combination of conflicting reactions. This calls for regulatory control that should be collocated at the 'junction' of the wirings. The lattice structure of Boolean closures provides natural junction

---

[2] The two canonical Boolean closures will be outlined in section 2.4

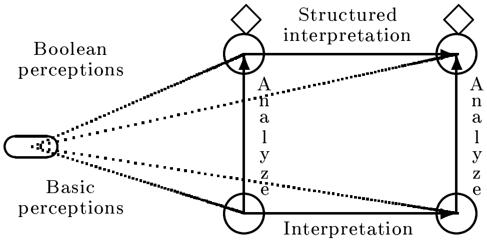

Figure 4: Two Boolean representations with interpretations

collocations for such wired control. Example reactions to conflicts: run away to avoid the conflict, or transit to a state of mind with no conflict ('sour grapes'), etc. (Being a non-trivial possession, integrative regulatory control could be lost, either because some types of emotions (e.g. romantic attraction in humans) may resist being subdued to regulatory control, or because systems may not be able to cope computationally with arbitrarily complex Boolean combinations of perceptual constituents. This could result in a (partial or total) derailing of attention and control.)

The generating arrow of figure 3 is a useful 'wall support'. If there exists a simple path (interpretation, communication) between two perceptions, as in figure 2, then this path is preserved also after the respective Boolean generations take place, and can be extended to a Boolean-structure-preserving-path between the generated representations. This property is mathematically warranted by the natural transformation. The underlying commutative diagram is shown in figure 4: A path from the lower left circle to the upper right diamond can be effected in either one of two possible ways: One could first generate a mental representation and then follow with a structured interpretation, or, alternatively, one could first follow a simple interpretation between basic perceptions, and then generate a structured mental representation that is based on the interpretation. This systematizes the interrelation between analytical and interpretive (communicative, learning) capabilities. (As stated by Barr and Wells (1995 [11]), commutative diagrams are *'the categorist's way of expressing equations'*.)

A salient property of the premises is the symmetry between $\mathcal{E}$, the environment, and $\mathcal{I}$, the representation. From a purely technical, context free, point of view, the roles that a w-element and a connotation play in the definitions are interchangeable. This *duality* has the technical consequence that any construct or theorem that is established for connotations (w-elements) can automatically be applied to w-elements (connotations), mutatis mutandis. The duality was applied to support a second wall that faces the wall from figure 4. A formalization of creative–imaginative processes is summarized in figure 5, which is dual to figure 4: it was technically based on mathematical results that were achieved by sweeping the roles of $\mathcal{E}$ and of $\mathcal{I}$. However, the cognitive processes that are formalized here are different. In perceptions with *conceived Boolean environments* the sets of w-elements are Boolean algebras, providing an adequate internal conception of combinations of similes and examples from the actual environment. (Boolean environments are designated here by an oval topped with a diamond.) This sets a basis for a planned perceptual–cognitive manipulation

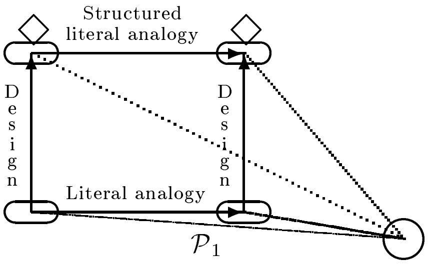

Figure 5: Conceived environments with analogies

of environments, for the creative imagination and rigorous planning of designs. Transitions between perceptions of conceived Boolean environments are based on Boolean homomorphisms of w-elements. They systematize structure aligning analogies.

Similar to the generation of mental representations, natural transformations formalized methodical cognitive transitions from authentic environments to conceived environments. A Boolean combination of w-elments is interpretable as a logical formula that can be further applied for a rigorous effective plan to realize the conceived design, marrying creative–imaginative capabilities with higher-level rational capabilities. The natural transformation warrants that, if there exists a simple analogy path between two environments, then this path is preserved by the respective Boolean generations, and can be extended to a Boolean-structure-preserving-path between the conceived environments. This is the import of the diagram in figure 5, that interrelates between analogies and creative design. A transition from the lower left oval to the upper right oval (with diamond) can be effected in either one of two ways: One could first conceive of a design and then follow with a structure aligning analogy to another design, or, alternatively, one could first follow a simple analogy between existing environments, and then conceive of a design that is already based on the analogical environment. Everything here pertains to a single set of connotations (the circle), and the discussion will be extended later.

Emotions that are conjured by perceptions of conceived environments systematize ‘what if’ emotions, perhaps like the deliberative layer in (Sloman, 2000 [25]). An agent that perceives an ulterior environment with its inner eye may feature emotive reactions ‘as if’ the imagined situation was real. Example emotions of this type could be anxieties that are caused by anticipation of failure or success that have not happened yet, but are internally conceived.

A composite diagram emerges: a base with two walls define a box, a whole that features more than the some of its parts. By figure 6, a ‘top cover’, two ‘side walls’, and two ‘diagonal walls’ are gained, representing more perceptions and composite transitions, all of which can be interwoven in a single architecture. The category theoretical equational reasoning affirms that the composite box commutes. Various AI cognitive habilitations are interrelated in a wider theoretical framework, with a high-level blueprint for an integrated computational framework.

Each one of the new walls describes a transition that takes a basic perception ($\mathcal{P}_1$ and $\mathcal{P}_2$, respectively) and scales it up to a cognitive perception with (ı) Analytic mental representation, (ıı) A perceptive inner eye that conceives of potential designs and plans, (ııı) Integrative emotional capabilities with autonomous regulatory control that are enabled by the Boolean structure as de-

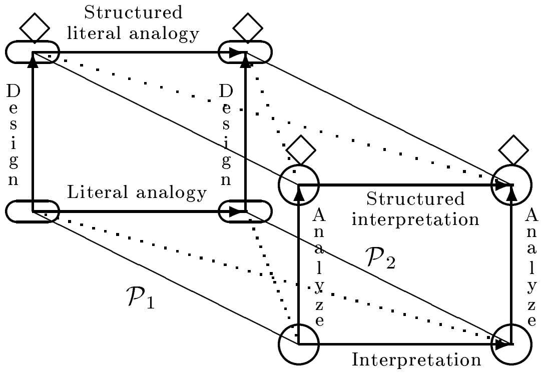

Figure 6: The synthesis

scribed before. The top cover describes an interpretive and analogical transition that applies Boolean homomorphisms to align the three high level capabilities that were just described.

Diagonals and diagonal walls of the diagram have to do with metaphorical perceptions, that were studied in (Arzi-Gonczarowski, 1999 [7]) (not all diagonals are shown in the figure). Emotions that are conjured by metaphorical perceptions may feature interesting discrepancies: A perceptual state that associates between an environment from one perception and connotations from another perception, could bring about emotive reactions that have developed relative to another *literal* context. They could be quite unexpected in the borrowed context.

2.3 Mental Activity

An agent could be initialized to a 'genetic inherited' perceptual state that features essential constituents: it attends to environmental chunks ($\mathcal{E}$) and to discriminations ($\mathcal{I}$) that are vitally consequential to its survival, and its urges and impulses are those that will make it endure. When the initial perceptual state lends itself to contingent transitions, adapts and matures, then perhaps it came with a certain 'mentality'. The various types of action tendencies that are formalized by the proposed schema are catalogued in (Arzi-Gonczarowski, 2000 [9]), with emphasis on motivations to actually perform the transitions that the diagram affords.

Besides transitions that happen as (either rationally planned, or instinctive) reactions to perceived constituents, some 'mind vitality' could also have its roots in action tendencies that are not related to the agent's relationship with its environment. Perseverant explorer types, for example, are often motivated by persistent drives. Behaviour could hence be driven by internal mental agendas as well. These are captured as built-in drives towards *attractor states* (although one may never really get to the attractor state)[3]. A formalization of such states is based on categorical 'terminal objects', that will be presented in the next section.

---

[3] A similar idea is offered by the dynamical systems stance in cognitive science.

2.4 Boundaries of the Schema

Mathematical tools that are afforded by the categorical formalism may be employed to systematize more intuitions about the confines of minds and intelligence. Whether the circuit box is bounded from various directions is the category theoretical version of questions regarding boundary conditions on equations.

2.4.1 Combinatorial Bounds

In the general case, p-morphisms add new constituents (the exceptions are mergers of synonym constituents). Hence, a simple type of bound that may be considered is on the number of different constituents. From the combinatorial point of view, the bound on the number $|\mathcal{I}|$ of connotations are $0 \leq |\mathcal{I}| \leq 2^{|\mathcal{E}|}$ for a given $\mathcal{E}$ (i.e. the possible subsets of w-elements circumscribe the discriminations that one may make). Dually, $0 \leq |\mathcal{E}| \leq 2^{|\mathcal{I}|}$ for a given $\mathcal{I}$ (i.e. the possible subsets of connotations circumscribe the distinct w-elements that one may conceive of). These are obvious bounds along the direction of the arrows.

The category theoretical version of stating that 'one cannot get any further than that' is to show that an object in a category (a perception in $\mathcal{P}rc$) is *terminal.* The *Total Universal Perception of* $\mathcal{E}$, $\mathcal{U}_{\mathcal{E}} = \langle \mathcal{E}, 2^{\mathcal{E}}, \epsilon \rangle$, with $2^{|\mathcal{E}|}$ connotations, has the existence property of (arrows leading to) a terminal object, and this lax[4] terminal object is unique up to isomorphism. This perception has the most evolved representation at the far right of the front wall of the box. Dually, a similar construct, with the *Universal Environment of* $\mathcal{I}$ that features $2^{|\mathcal{I}|}$ w-elements, has the existence property of (arrows leading to) a terminal object, and this lax terminal object is unique up to isomorphism. This perception has the most evolved conceived environment at the far right of the back wall of the box. These boundary perceptions marry the combinatorial aspect with the categorical algebraic language. They are the *attractor states* mentioned in the former section.

The initial object for the category is the *Empty Perception* $\mathcal{P}_{\emptyset} = \langle \emptyset, \emptyset, \varrho_{\emptyset} \rangle$. It stands for 'no environment and no representation', and puts a theoretical bound on the 'origin' of the arrows, from the left and from the bottom of the box (perhaps a theoretical *tabula rasa*).

2.4.2 A Fixed Point Bound

A stronger result than the above was also shown, deploying the strengths of the categorcal setting to systematize more intuitions about intelligence. Figuratively, the 'top cover' of the 'box' could perhaps serve as a base for another box, and the question is whether it is possible to 'pile up' infinitely many boxes, one on top of the other. This would have meant that a mind could infinitely improve its high level capabilities, constantly adding more compound concepts, more plans and designs, and more integrated behaviours.

The vertical arrows of the diagrams are based on perception endofunctors of the form $\mathcal{G} : \mathcal{P}rc \rightarrow \mathcal{P}rc$, where $\mathcal{G}(\mathcal{P})$ is a Boolean perception. A vertical arrow $\xi : \mathcal{P} \rightarrow \mathcal{G}(\mathcal{P})$ is a natural transformation from the identity functor on $\mathcal{P}rc$ to

---

[4] The uniqueness property of (arrows leading to) this perception does not hold in the 3-valued context.

the functor $\mathcal{G}$. If $(\mathcal{P}, h)$ were a *fixed point* of $\mathcal{G}$, then $\mathcal{G}(\mathcal{P})$ would be the same as $\mathcal{P}$, practically stopping the 'piling up of boxes'. This would mean that $(\imath)$ The cognitive transition that is systematized by $\mathcal{G}$ is unable to further scale up perception beyond that which is already featured by $\mathcal{G}(\mathcal{P})$. $(\imath\imath)$ $\mathcal{G}$ is a sensible cognitive process that knows its limitations and is 'aware' of property $(\imath)$.

Two canonical Boolean closures were studied in (Arzi-Gonczarowski and Lehmann, 1998 [3], Arzi-Gonczarowski, 1999 [6]). Only one of them features a fixed point (Arzi-Gonczarowski, 2000 [8]). The difference between them is related to validity and completeness in Boolean perceptions. These notions are based on relationships between the Boolean partial order $\leq$ on constituents (connotations, w-elements) on one hand, and perceived lawlike patterns on the other hand. Formally, the perceptual quasi order $\trianglelefteq$ is defined: $(\imath)$ For $\alpha, \beta \in \mathcal{I}$, $\alpha \trianglelefteq \beta$ if $\forall w \in \mathcal{E} \quad \varrho(w, \alpha) = t \Rightarrow \varrho(w, \beta) = t$ and also $\varrho(w, \beta) = f \Rightarrow \varrho(w, \alpha) = f$. $(\imath\imath)$ For $x, y \in \mathcal{E}$, $x \trianglelefteq y$ is defined in a dual manner. Lawlike patterns may also involve Boolean combinations of constituents.

As already explained in section 2.2, since Boolean lattices feature a partial order, this enables the organization of connotations in hierarchies. In a *valid* Boolean perception $\leq \subseteq \trianglelefteq$, meaning that the formal Boolean hierarchy can be verified by perceptual observations. In a *complete* Boolean perception $\trianglelefteq \subseteq \leq$, meaning that all observed lawlike patterns are reflected in the Boolean structure. Boolean perceptions are always valid, but not necessarily complete. Perceptions in the *valid and complete* Boolean subcategory, $\mathcal{P}rc^{\text{bl-cmp}}$, feature total internalization of perceived lawlike patterns[5].

The simplest Boolean closure takes the constituents of basic perception as free generators, defines a free functor $\mathcal{G}^{\text{fr}} : \mathcal{P}rc \rightarrow \mathcal{P}rc^{\text{bl}}$, and systematizes a general cognitive transition from basic perceptions to Boolean perceptions. It captures methodicalness and open-mindedness, but not perceptual acuity, because $(\imath)$ $\mathcal{G}^{\text{fr}}(\mathcal{P})$ is, in the general case, incomplete (freedom means that there is no dependence between constituents, which is the essence of lawlike patterns). $(\imath\imath)$ $\mathcal{G}^{\text{fr}}$ has no fixed point. In particular, $\mathcal{G}^{\text{fr}}$ is unable to 'sense' a case where $\mathcal{P}$ is already a Boolean perception, and it unconditionally generates a Boolean set of 2^{2^n} constituents over any given n constituents. (A combinatorial explosion will be avoided when the 'pile' eventually hits the general combinatorial upper bound of section 2.4.1).

The sketch-structure of perceptions (Arzi-Gonczarowski and Lehmann, 1998 [3]) answers the imperviousness of $\mathcal{G}^{\text{fr}}$. Loosely, a p-morphism in the sketch-structured subcategory, $\mathcal{P}rc^{\text{Sk}}$, preserves lawlike patterns, namely the quasi order $\trianglelefteq$ (the technical details can be found in the cited works). The endofunctor $\mathcal{G}^{\text{fr-cmp}} : \mathcal{P}rc^{\text{Sk}} \rightarrow \mathcal{P}rc^{\text{bl-cmp}}$ is a free functor. Loosely, it 'moves things around' in the Boolean lattice to reflect the perceived patterns. Consequently, the transition is perceptually acute: $(\imath)$ $\mathcal{G}^{\text{fr-cmp}}(\mathcal{P})$ is valid and complete: it features total observation and internalization of all lawlike patterns that are perceptible by $\mathcal{P}$. $(\imath\imath)$ For all valid and complete Boolean perceptions $\mathcal{P}$, $(\mathcal{P}, \xi^{-1})$ is a fixed point of $\mathcal{G}^{\text{fr-cmp}}$. This is a sensible cognitive process that knows its limitations, it is 'aware' of property $(\imath)$, and would not modify perceptions that it is unable to amend.

The fixed point formalism tells us that $\mathcal{G}^{\text{fr-cmp}}$ is superior to $\mathcal{G}^{\text{fr}}$, not only

---

[5] Detection of lawlike patterns can be based on a programmed implementation like LAD (Boros and al., 1996 [13]).

because it is more perceptually acute, but also because it has an 'awareness' that avoids the 'unnecessary piling up of boxes'. This bound is cognitively derived from within, on the basis of own observations and own intelligence. This is different from the 'bureaucratic' combinatorial bound that has nothing to do with innate perceptual capabilities. Familiar intuitions that have just been systematized are ($\imath$) Abstract speculations are not enough for real cognizance. A perceptive agent should acutely relate to its environment to construct a truly intelligent mental representation. ($\imath\imath$) Sensible cognitive processes should be aware of their limitations.

Based on the observation that the category of $\mathcal{G}^{\mathrm{fr-cmp}}$–algebras is, in particular, a generalized poset, one gets a hierarchy of valid and complete Boolean perceptions as fixed points $(\mathcal{P}, \xi^{-1})$ of $\mathcal{G}^{\mathrm{fr-cmp}}$. (Figuratively: a sequence of bounded walls of ascending size.) This systematizes the intuition that among perceptions with equally advanced Boolean capabilities (namely $\mathcal{G}^{\mathrm{fr-cmp}}$), those with the more detailed grounding apparatus will generate better cognizance. The initial, empty, perception makes the *least fixed point* (a zero size wall). This captures the intuition that even with the best speculative mind, no true cognition can emerge if there is no grounding apparatus that interacts with an authentic environment. Cognition is both enabled and circumscribed by perception.

2.5 So Far

Like a reduced instruction set for a computer, the formal ontology that has been outlined conflates the types of building blocks that are required for a high level architecture (w-elements, connotations, p-predicate, categorical primitives, Boolean primitives), but not necessarily the spectrum of mind mechanisms that are modelled. The next step is to show how the same building blocks can be deployed to formalize aspects of consciousness.

3 Perceptions that perceive themselves

3.1 Senses of self

Let $\mathcal{P} = \langle \mathcal{E}, \mathcal{I}, \varrho \rangle$ be a perception as in section 2. From the technical point of view, there seems to be no problem with a 'recursive' call: $\varrho([\mathcal{P}], \alpha)$. As a matter of fact, why not consider also:

$$\varrho([\mathcal{P}'], \alpha) \;\;,\;\; \varrho([\mathcal{P}' \to \mathcal{P}], \alpha) \;\;,\;\; \varrho([\mathcal{P} \to \mathcal{P}' \to \mathcal{P}''], \alpha) \;\;,\;\; \ldots$$

One could proceed to perceive various $\mathcal{P}rc$ categorical constructs and diagrams[6]. For instance, one could 'squeeze' (an instance of) the box of diagram 6 into one of the ovals, making it a w-element in, say, $\mathcal{E}_1$. Of course, proper care has to be taken to avoid confusions. For example, when a p-morphism h is applied, then $h : \mathcal{P} \to h(\mathcal{P})$ is not the same as the element mapping $h : [\mathcal{P}] \mapsto h([\mathcal{P}])$. Perceptions like the above, that are capable of perceiving (instantiated) entities from the proposed schema, will be called *reflective perceptions.* Connotations

---

[6] The square brackets are not really necessary, but they may help the distinction of 'local' variables.

and emotive reactions that are associated with *reflective w-elements* will be called *reflective connotations*, *reflective reactions*, etc.

It should be noted that, from the theoretical point of view, reflective constituents challenge the *iterative hierarchy* of the proposed constructs: begin with some primitive elements, then form all possible perceptions with them, then form all possible perceptions with constituents formed so far, and so on. In set theoretic notions this has to do with a possible violation of the *axiom of foundation*, normally added to the five original axioms of Zermelo, that warrants against some set theoretic paradoxes. Hence, $\mathcal{E}$ and $\mathcal{I}$ may not be classical sets when they contain reflective elements. Following Dubois (1998 [16]), readers are refered to discussions on these issues by Aczel (1987 [1]) and by Barwise and Moss (1991 [12]).

Some examples of 'senses' (Damasio (1999 [15]) sometimes calls them *super*senses) that reflective perceptions could formalize:

1. A sense of self and the boundaries of the self could be systematized using ownership connotations such as:
$\varrho([\mathcal{P}], \mathit{my\_perception}) = t$, or $\varrho([\mathcal{P}'], \mathit{my\_perception}) = f$.

2. A sense of the perceiving self could be systematized as follows: Let the reflective w-element be $[\mathcal{P}]$, and the reflective connotation be $[\varrho(w, \alpha) = x]$. The value of the reflective p-predicate $\varrho([\mathcal{P}], [\varrho(w, \alpha) = x]) = y$ then means:

> It is y (namely: t or f or u) that perception $\mathcal{P}$ has the property that it perceives the w-element w as having/lacking/... connotation α.

If $y = t$ had invariably held when, and only when, $\varrho(w, \alpha) = x$, then the discrimination might not have been so interesting. However, we know that even healthy humans sometimes fail to make perceptual reports, on occasions we need to pinch ourselves to make certain that we are not hallucinating, and on other occasions we may be unaware of a noise, an itch, etc.

Perceptions and emotions are tightly coupled, hence this experiential feel is also tightly coupled with the next example:

3. A sense of the emoting self could be systematized as follows: Let the reflective connotation be $[\varrho(w, \alpha) = \mathit{x\_triggers\_Z}]$, namely 'The combination of α, w, and $\varrho(w, \alpha)$ activates a reaction Z'. If a call to the reflective p-predicate $\varrho([\mathcal{P}], [\varrho(w, \alpha) = \mathit{x\_triggers\_Z}])$, returns y, then this means:

> It is y (namely: t or f or u) that perception $\mathcal{P}$ has the property that its emotive reaction is Z to the perception of the w-element w having/lacking/... connotation α.

Again, even healthy humans may react absent-mindedly[7], neurologists report about episodes of automatism, and quite a few psychoanalysts would have been out of work if such reflections had been trivial. (One may actually never directly examine one's own gaze, even in the mirror, without effecting that gaze itself. That perhaps explains why we are hardly ever aware of our own eye movements, that disclose more about us than that which we would usually be willing to admit.)

The example formalizations above seem to capture a phenomenon that Damasio calls *core consciousness*, which provides the agent with a sense of its *core self* about one moment - now - and about one place - here - given that $\mathcal{P}$ is the

---

[7] Behaviour without report is necessary for some everyday skills (see section 3.5).

current perception. Damasio (1999 [15, p.26]) says: '*… consciousness begins as the feeling of what happens when we see or hear or touch … the feeling marks those images as ours and allows us to say … that we see or hear or touch.*

The following examples seem to formalize a more complex kind of consciousness that Damasio calls *extended consciousness*, which provides the agent with an elaborate sense of self, an identity, and places that identity at a point in individual historical time. These examples involve perceptions of structures that are more complex than just the current perceptual state $\mathcal{P}$. They require memory and more symbolic representations:

4. A sense of the transition of time could be formalized by:
$\varrho([\mathcal{P}'], yesterday's\_perception) = t$, or
$\varrho([\mathcal{P}'], today's\_perception) = f$, or
$\varrho([\mathcal{P}' \to \mathcal{P}], transition\_from\_yesterday) = t$.

5. Other senses of the transient self could also be formalized by perceptions of p-morphisms as w-elements. Reflective connotations may then make discriminations about them, classifying them as transitions of location, or interpretive transitions, as well as all the other types of transitions that have been discussed in section 2. Compositions of multiple arrows may stand for longer biographies.

As explained in section 2, the schema systematizes a feature that specific p-predicate values typically conjure emotive reactions. Some example formalizations of reflective emotive reactions could be:

1. Sensing the emoting self, as described above (namely: It is $t/f/u$ that perception $\mathcal{P}$ has the property that its emotive reaction is Z to the perception of the w-element w having/lacking/… connotation α), may itself conjure an emotion. This formalization of 'the feeling of an emotion' is suitable to capture conscious aspects of affect. Just for survival, it would have probably been enough to flee danger like a zombie, without perceiving oneself in the act of flight. The reflective capability conjures the conscious feeling of, say, fear. It is formalized by the (second order) reaction to the sense of the emoting self. Similarly, it would have probably been enough to eat or reproduce like a zombie, without the reflective sense of self in the act of emoting. It is the reflective capability that conjures the conscious feeling of, say, pleasure. The significance of reflective emotions to intelligent behaviour is discussed below.

2. When an agent senses its perceiving and emoting self, then a reflective emotive reaction could perhaps consist of the tendency to change the state of things. Reflective emotions are capable of systematizing that. A drive to modify the state of things could encourage creativity (arrows for that have been described in (Arzi-Gonczarowski 1999 [6])). In other cases reflective emotive reactions open the possibility to capture self criticism and other conscious emotions: When an agent senses its emoting self, it may perhaps react to manipulate its own attitude, for instance by suppressing, hiding, or amplifying the emotion. In the human context, example cases are self critical emotions such as shame or pride. (Emotive reactions are not necessarily fully realized: an emotion is a *tendency* to react, so some, or all, of the above may not be overt.)

From the evolutionary point of view, survival pressures are behind emotions, and even more so behind reflective emotions and consciousness: Emotions are about reacting to the environment in a way that happened to support the general preservation of the species. Reflective emotions allow a purposeful management and optimal monitoring of these reactions to the advantage of the individual self. Zombies may have automatic 'one size fits all' emotive reactions, but only conscious perceptions are able to tailor specific successful behaviours with individually controlled forethought, flexibility, and creativity. That is their competitive edge (see (Ballonoff, 2000 [10])). They also have drawbacks, some of which will be discussed in section 3.5 below.

3.2 Anticipation

When an agent has the tools to perceive perceptions, there is no reason why it should restrict this capability to just its own current perceptual state. Some examples from the former section (reflective perceptions of p-morphisms) already had to do with former perceptual states as well. This possibility can be further explored to formalize more mental capabilities, systematizing the intuition that they are dependent on consciousness as well.

Conceived perceptual states have been introduced in section 2.2. They are about environments and situations that are created internally, and are based on similes from authentic perceptual experiences. This type of cognitive processes is formalized by p-morphisms from authentic perceptual states to conceived perceptual states. That is how one may go about formalizing agents that plan, create (and also avoid unwanted) situations, just by evaluating them internally. In reflective perceptions of arrows, $\varrho([\mathcal{P} \to \mathcal{P}'], \alpha)$, $\mathcal{P}'$ may hence also be a conceived perceptual state that is anticipated in the future. Damasio calls them *memories of the future*, because the internal images that make them are essentially similar to those that pertain to memories of the past. For example, one may create an internal image of what happened yesterday, and one may also create an internal image of things that are anticipated to happen tomorrow, and both may be vividly conceived and just as consequential to one's current affective and cognitive, overt and introvert, behaviour. Damasio's notion of extended consciousness is about a sense of an *autobiographical self* that is aware not only of the lived past but also of the anticipated future, where they are *'sensed along with the here and now in a sweeping vista as far-ranging as that of an epic novel'* (Damasio, 1999 [15, p.17]). It is proposed here that Damasio's autobiographical sense of self and extended consciousness may be systematized by reflective perceptions of categorical compositions of multiple arrows, rigorously modelling a path from remembered past states to conceived future states:

$$\varrho(\ [\ \ldots\ \to \mathcal{P}_{now-1} \to \mathcal{P}_{now} \to \mathcal{P}_{now+1} \to\ \ \ldots\ ]\ ,\ \mathit{This\_is\_me}\ )$$

Dubois (1998 [16]) defines an anticipatory system as a system which contains a model of itself and/or of its environment, and computes its present state as a function of the prediction of the model. In (2000 [17]) he discriminates between weak anticipation where future states are just a recall of the past states projected to the future, and strong anticipation which is built by the system itself in an incursive way. Using this terminology, a reaction to a reflective perception such as the formula above, would be a case of *incursion* rather than recursion, because the system then computes its behaviour taking into account not only past or present perceptual states, but also future perceptual states. Eventual connections between this research and research on anticipatory systems will be further explored in section 3.5 below.

3.3 Empathy and Creativity

Having accepted that the tools to perceive perceptions should not necessarily be restricted to own perceptual situations that are here and now, let us formalize perceiving *perceptions of other agents*, and a systematization of how they relate to consciousness will naturally follow.

A perception $\mathcal{P}'$ may indeed be a w-element in the environment $\mathcal{E}$ of another perception $\mathcal{P} = \langle \mathcal{E}, \mathcal{I}, \varrho \rangle$. Example relevant connotations are ownership connotations that connote the boundaries of the self, as in section 3.1. As a matter of fact, this is not really a case of a reflective constituent, because $\mathcal{P}'$ is external to $\mathcal{P}$. The reflective perspective comes from analogies, metaphors, and creative conceptions through similes. With the various cognitive transitions that are outlined in section 2 and formalized by p-morphisms, one may *conceive of oneself as being in the perceptual state* $\mathcal{P}'$. In that conceived perceptual state, $\varrho([\mathcal{P}' \wedge \{me\}], \alpha)$ is a reflective call to the p-predicate[8]. An emotive reaction to the conceived situation may conjure empathy: participation in the experience of the other agent's feelings. This is a formal modelling of how agents may empathize with (and also manipulate) others. Damasio (1999 [15, p.5]) describes empathizing with a frail old man who is having difficulty embarking on a ferry: '*... without consciousness, the old men's discomfort, perhaps humiliation, would simply not have been known to him. Without consciousness, the two men on deck would not have responded with empathy. Without consciousness, I would not have been concerned and would never have thought that one day I might be him...*'.

When applied to categorical compositions of multiple arrows, conceived reflective perceptions could perhaps formalize creative processes similar to those of novelists and playwrightes, who are capable of vividly conceiving of 'biographical selves' other than their authentic own ones.

From the evolutionary point of view, social pressures are probably behind the exaptation of core conscious capabilities to empathy (and its darker side: manipulativity).

3.4 Phenomenal Grounding

Perceptions as defined in section 2 formalize embodied perceptual states. For a perception $\mathcal{P} = \langle \mathcal{E}, \mathcal{I}, \varrho \rangle$, the value of $\varrho(w, \alpha)$ is conceived to be phenomenally grounded by the sensory motor neural apparatus of that agent, namely its physical embodiment. In simple words, the agent should *really* experience w as having/lacking/... the connotation α. Of course, things may be emulated, but then one enters into Searle's Chinese room (Searle, 1984 [24]). For humans, science knows quite a lot about the biological underpinnings of the classical senses: smell, touch, taste, hearing, and vision. AI has been able to endow artificial agents with some unmediated sensory motor neural capabilities as well.

In the case of reflective perceptions one should not settle for less, namely that perceptions should perceive themselves phenomenally, physically experiencing themselves and their perceptual affective and cognitive processes from a 'first person' perspective. Science is still struggling to figure out the biological underpinnings of natural consciousness. Damasio (1999 [15]) conjectures that the

---

[8] (Arzi-Gonczarowski, 1999 [6]) shows how to conceive of w-elements applying sets of connotations and set operations.

roots for the self are found in the *proto self*, the ensemble of brain devices which continuously perform the automated maintenance of the body state within the narrow range and relative stability required for survival. The uninterrupted continuity from birth to death is crucial, and also the innate widespread interconnections of function: it is the same brain that perceives outwards and inwards. It is noted here that the proto self is different from the superficial internal scanning that is performed by ordinary computers, which is done from a 'third person' perspective, is not necessarily continuous, and is not so crucially and autonomously responsible for survival.

The proposed schema does not offer to solve the problem of the phenomenal grounding of self perception. It is a high level structural proposal, and substitution instances will need to be phenomenally grounded to actually have reflective capabilities.

3.5 The Catch: Wasted Resources

What about reiteration, namely perceptions that perceive themselves perceive themselves perceive themselves. . . ? One may inquire:
1. Whether this is necessary or interesting for agents to do that.
2. What natural capabilities, if at all, are captured by more levels of reflection that cannot be captured otherwise.
3. What if an agent gets into an infinite recursion, requiring more and more resources, eventually derailing the system.
4. Whether there is a halting condition or a theoretical fixed point that could warrant against infinite regress.

These issues have not been yet looked into in the context of the proposed schema, but one may try to conjecture on the basis of related research, because attention to the connection between intensive conscious reflection and the danger of derailed behaviour has been drawn before. Some everyday evidence comes from the fact that the acquiring of skills such as car driving, sports playing, music making, and the like, involves the internalization of basic automatic behaviour in order to save time. Resource consuming, 'expensive', conscious reactions are spared for the creative aspects of these activities, or for extreme situations. Marvin Minsky has been heard saying that *consciousness is what you do when things go wrong.* Sloman (2000 [25]) 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'.* Flanagan (1998 [18]) suggests that the role of consciousness may be regarded as '*. . . more one of interfering with cognitive processes that are designed to function well and generally do so. . .* '.

A possible solution has been suggested by Dieter Gernert, that specific cases, where n nesting reflections are practically necessary, are probably rarer as n grows, and that is perhaps how reflective efforts could converge.

Even more optimistic conjectures could be based on research about anticipatory systems with incursion. As already mentioned in section 3.2, Dubois (1998 [16], 2000 [17]) defines an anticipatory system as a system which contains a model of itself and/or of its environment, and computes its present state as a function of the prediction of the model:

$$x_{t+1} = F(\ \ldots\ ,\ x_{t-1}\ ,\ x_t\ ,\ x_{t+1}\ ,\ \ldots\ )$$

x_{t+1} on the right hand side of the equation can be replaced indefinitely by the entire equation. Dubois has shown that the infinite incursive equation

$$x_{t+1} = 4\ x_t\ (1 - x_{t+1})$$

converges to the following recursive equation $x_{t+1} = 4x_t/(1 + 4x_t)$, for which a stable state solution exists. When there is no anticipation, $x_{t+1} = 4\ x_t\ (1 - x_t)$, the system is chaotic with an unstable state solution. Incursion transforms an unstable state of the system to a stable state, so without anticipation, such a system cannot converge to a stable solution. In the chaotic system, the system never reaches a stable state and its behaviour is sensitive to the values of the initial conditions and becomes unpredictable. In the incursive equation, the system reaches a unique stable state for any values of the initial conditions. With anticipation, the chaotic system becomes an incursive system which is independent of initial conditions to reach a stable final state, a fixed point which is the predictive objective of the anticipation. Dubois discusses a rich variety of 'system selves' that can be described by incursion, avoiding infinite regress, such as models of population dynamics, growth of bacteria, biochemical reactions, and quite a few physical systems.

The x's above could perhaps translate to perceptual states of this paper, as proposed in section 3.2:

$$\varrho(\ [\ \ldots\ \to \mathcal{P}_{now-1} \to \mathcal{P}_{now} \to \mathcal{P}_{now+1} \to\ \ \ldots\ ]\ ,\ \alpha\ )$$

These perceptions may be associated with reactions that conjure actual perceptual affective and cognitive transitions, $\mathcal{P}_{now} \to \mathcal{P}_{now+1}$, yielding the next perceptual cognitive and affective state $\mathcal{P}_{now+1}$. An interesting goal would be to instantiate the perceptual structures of this paper by systems that can be described by incursion, as analyzed by Dubois. A general reduction, if it exists, is unlikely to be a simple one, because the $\mathcal{P}$'s are not uniform entities neatly formulated by numerical or vectorial variables and equations. The $\mathcal{P}$'s are made up of heterogenous collections of:

1. Environmental elements not limited to objects or events of any type,
2. Internal discriminations not limited to symbols or diagrams or procedures or formulas, or any other representational format,
3. Behaviours not limited to physical, or mental, or social, or any other essential nature.

. . . and so on. ISAAC provides a unifying formal framework for that open ended diversity by virtue of a categorical abstraction. Minsky (2000 [22, p.71]) expresses pessimism about such reductions: *'Physics researchers have made great progress by searching for elegant unified theories. But* AI *must deal with different complex worlds than the ones theoretical physicists face because they must deal with specific things that emerged from more inhomogeneous processes. We cannot expect to find uniform explanations to deal with that much diversity. Instead, we'll have to invent, combine, and reorganize an ever-increasing collection of increasingly incomplete theories.'*

In humans, infinite regress is also avoided because they are naturally unwilling or reluctant to infinitely reflect and wonder about themselves. A formal version of that may perhaps turn out to be a certain 'innate mental laziness' (not imposed from outside), preserving resources or redirecting attention towards more diverting[9], less wearying, issues.

4 Methodological Fallout

One of the reasons that mathematical modelling often contributes to scientific understanding, is that the mathematization enables us to pinpoint, meticulously and rigorously, the premises that are being applied at every stage (see, for example, (Croon and de Vijver, 1994 [14])). By the reasoning and the constructions of this model, one may notice that:

1. The basic constructs of the schema, from section 2.1 (w-elements, connotations, p-predicate, categorical primitives), are sufficient to model reflective perceptions. It is the reiterated call to the p-predicate that 'does the job'. Higher level constructs of the schema (of section 2.2) like functors and natural transfomations, and the Boolean constructs (that are associated with analytic reasoning and related issues) are not required to formalize reflective capabilities. In this respect the formalization supports researchers, like Lycan (1996 [20]), who argue that there is no more to consciousness than can be accounted for in terms of intentionality, functional organization, and, in particular, second-order representation of one's own mental states.
2. However, the integration of the higher level constructs (of section 2.2) with reflective capabilities is required to formalize the best of conscious intelligence: anticipation, creativity, empathy...
3. The formalization supports intuitions that some intelligent capabilities, like empathy and fictional narration, which are not directly about the perceiving self, are nonetheless tightly connected to consciousness (section 3.3).
4. Reflective capabilities, as formalized by the proposed schema, provide representations with perceptual cognitive and affective content. The associated reflective w-elements, connotations, and emotive reactions, can then be mapped and manipulated by p-morphisms and other categorical constructs. This provides formal basis for *access consciousness*: that the content of consciousness is available as a premise in reasoning and other mental activities, and lends itself to rational control of action and of symbolic representation.

5 Conclusion

ISAAC, the integrated schema of affective artificial cognition, is capable of formalizing aspects of consciousness through recursive calls of the perception procedure to itself. Among the mental capabilities that are systematized are various senses of self, senses of time and other transitions, reflective emotional capabilities, anticipation, empathy and creativity.

References

[1] Aczel P. (1987) *Lectures in Nonwellfounded Sets.* CSLI Lecture notes 9. CSLI.

---

[9] Consider how *diversion* acquired the meaning of amusement or entertainment.

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

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

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

[5] Arzi-Gonczarowski Z. (1998) Wisely non rational – a categorical view of emotional cognitive artificial perceptions. *Papers from the 1998 AAAI Fall Symposium: Emotional and Intelligent: The Tangled Knot of Cognition.* Edited by D. Cañamero, published by AAAI, pp. 7–12.

[6] Arzi-Gonczarowski Z. (1999) Categorical tools for perceptive design: Formalizing the artificial inner eye. *Computational Models of Creative Design IV.* Edited by J.S. Gero and M.L. Maher, published by the Key Centre of Design Computing and Cognition, University of Sydney, Australia, pp. 321–354.

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

[8] Arzi-Gonczarowski Z. (2000) A blueprint for a mind by a categorical commutative diagram. *Proceedings of the AISB'00 Symposium on How to Design a Functioning Mind.* Published by the Society for the Study of Artificial Intelligence and the Simulation of Behaviour, UK, pp. 10–18.

[9] Arzi-Gonczarowski Z. (2000) A categorization of autonomous action tendencies: The mathematics of emotions. *Cybernetics and Systems 2000.* Published by the Austrian Society for Cybernetic Studies, Vienna, vol. 2 pp. 683–688.

[10] Ballonoff P. (2000) On the evolution of self-awareness. *Cybernetics and Systems 2000.* Published by the Austrian Society for Cybernetic Studies, Vienna, vol. 1 pp. 347–352.

[11] Barr M. and Wells C. (1995) *Category Theory for Computing Science.* Prentice Hall.

[12] Barwise J. and Moss L. (1991) Hypersets. *The Mathematical Intelligencer* 13(4), pp. 31–41.

[13] 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. Published by rutgers University, New Brunswick, N.J.

[14] Croon M.A. and Van de Vijver F.J.R. (editors) (1994) *Viability of Mathematical Models in the Social and Behavioral Sciences.* Swets and Zeitlinger B.V., Lisse.

[15] Damasio A.R. (1999) *The Feeling of What Happens.* Harcourt Brace & Company.

[16] Dubois D.M. (1998) Computing anticipatory systems with incursion and hyperincursion. *Proceedings of CASYS, First International Conference on Computing Anticipatory Systems.* Edited by D.M. Dubois, published by the American Institute of Physics, AIP Conference Proceedings 437, pp. 3–30.

[17] Dubois D.M. (2000) Review of incursive, hyperincursive and anticipatory systems – foundation of anticipation in electromagnetism. *Proceedings of CASYS'99, Third International Conference on Computing Anticipatory Systems.* Edited by D.M. Dubois, published by the American Institute of Physics, AIP Conference Proceedings 517, pp. 3–30.

[18] Flanagan O. (1998) Consciousness. *A Companion to Cognitive Science.* Blackwell, chapter 9 pp. 176–185.

[19] Gould S.J. and Vrba E. (1982) Exaptation: A missing term in evolutionary theory. *Paleobiology* 8, pp. 4–15.

[20] Lycan W.G. (1996) *Consciousness and Experience.* MIT Press.

[21] Magnan F. and Reyes G.E. (1994) Category theory as a conceptual tool in the study of cognition. *The Logical Foundations of Cognition.* Edited by J. Macnamara and G.E. Reyes, published by Oxford University Press, pp. 57–90.

[22] Minsky M. (2000) Commonsense-based interfaces. *Communications of the ACM* 43(8), pp. 67 – 73.

[23] Seager W. (1999) *Theories of Consciousness.* Routledge.

[24] Searle J.R. (1984) *Minds, Brains, and Action.* Harvard University Press.

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

[26] Sloman A. (editor) (2000) *Proceedings of the AISB'00 Symposium on How to Design a Functioning Mind.* Published by The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, UK.