Understanding One Another:
Making Out Meanings with Boolean Equations

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

April 3, 2007

Abstract

Autonomous AI agents raise the issue of semantic interoperability between independently architectured and differently embodied intelligences. This article offers an approach to the issue with certain aspects that are close in spirit to the way humans make out meanings. Using a mathematical model of cognition, it is shown how agents with autonomously developed conceptualizations can bootstrap and unravel each other's meanings ad hoc. The domain general methodology is based on the agents' capability to deal with Boolean operations, and on the shared outside environment. No prior provisions are required. The formalized cognitive process consists of constructing, and solving, Boolean equations that are grounded in the shared environment. The process yields a testable conjecture about the grounded conceptual representation of the other, along with a testable conjectured translation that maps from that representation to one's own.

1 Introduction

Autonomous AI agents raise the issue of semantic interoperability between independently architectured and differently embodied intelligences. When the infrastructure for intelligence emerges bottom-up from its embodiment and an own sensory-motor-neural apparatus, then the way each agent perceives the environment, and the ensuing conceptualizations and ontologies, could vary radically.

Autonomous agents cannot 'look into each other's head'.(Researchers of consciousness often use the term *privileged access* in relation to phenomena that are not overt to others.) A typical strategy of creating large, shared, conceptual schemas to be used as a global common reference is problematic, because they are not grounded in the individual perceptions of the agents that access them. That would require pretailored specific translations to a fixed conceptualization of a database. In a general case it cannot be anticipated that intelligences that are independently architectured should eventually need to communicate. It is hence neither possible to negotiate and correlate meanings while they are emerging, nor to prepare the mechanisms required for a shared reference.

This article offers an approach to the issue that is more adaptive, and closer in spirit to the way humans make out meanings when confronted with a particular necessity to do so. (Closeness in spirit does not mean similarity in all internal procedural details, but rather in certain apparent behavioral aspects.) The approach is based on having communicating intelligences bootstrap and unravel meanings ad hoc, using an intelligent cognitive apparatus, but no specific anticipation, provisions, or protocol. This is formalized by an autonomous construction and solution of grounded Boolean equations.

The study is operated within the general framework of ISAAC (*I*ntegrated *S*chema for *A*utonomous *A*ffective *C*ognition), which is a methodologically oriented, long term research that uses algebraic and categorical formalisms with the goal of setting the modeling of autonomous intelligent agents on a unified and rigorous mathematical basis. A category has been defined where every intelligent state, including its conceptualizations, is a categorical object, and morphisms are paths of commensuration between these states. Different intelligent states could pertain either to different agents, or to the same agent in different settings/stages of its development, hence morphisms could be either inter- or intra- agent. (The whole framework will be further explained in section 3.)

With ISAAC's formalism, the first contribution of this work to the problem at hand, even before the solution is provided, is by casting the problem in a rigorous form of a well defined mathematical question: Given two objects in the given category, one is looking to define a morphism between them. The morphism should satisfy specific constraints that capture grounded meaning preservation. The obtained morphism would model a meaningful translation between the modeled intelligent states and their conceptualizations.

When questions can be formally stated, and they will also be answered here with mathematical tools, theorems, and proofs, there is of course an added value of rigor and generality that is typically gained with mathematical modeling. This is the main advantage of ISAAC in general, and of the contributions of this article in particular.

Section 2 provides general background on the problem and contextualization within AI.

Section 3 outlines the underlying formalism of ISAAC, to make the presentation self contained. It summarizes foregoing constructions, results, and related interpretations of ISAAC's research agenda, that have already been published and presented. This is necessary for the understanding of section 4 that follows.

Section 4 details and discusses this article's novel contribution about a mathematically modeled technique to autonomously make out meanings, showing how it falls naturally out of ISAAC's formalism.

2 General Constextualization

2.1 Moments in the History of an Age-Old Problem

The problem that is addressed in this article is about a suitable translation between the meanings of two different conceptualizations, for purposes of intelligent communication. This problem has been of interest to research communities in linguistics, philosophy, game theory, economics, cognitive science, machine translation, and AI. Within each one of these communities the problem is of course instantiated in a slightly different form.

The significance of communicating meanings 'correctly' has probably been appreciated in the earliest societies. For example, this is the main issue in one of the first biblical stories - the story about the tower of Babylon: *'And the Lord came down to see the city and the tower, which the children of men were building. And the Lord said, Behold, the people is one, and they have all one language; and this they begin to do: and now nothing will be withheld from them, which they have schemed to do. Come, let us go down, and there confound their language, that they may not understand one another's speech. So the Lord scattered them abroad from there upon the face of all the earth: and they ceased to build the city. Therefore is the name of it called Bavel*[1]*; because the Lord did there confound the language of all the earth: and from thence did the Lord scatter them abroad upon the face of all the earth.'* (Genesis 11,5-9)

A momentous event in the modern history of instances of the problem was the 19th century deciphering of the inscription on the Rosetta Stone, leading to the understanding of hieroglyphic writing. The Rosetta Stone (now in the British Museum) was written in Egypt about two centuries BC. It is inscribed in two languages, Egyptian and Greek, and three alphabets: hieroglyphic, demotic (a cursive form of Egyptian hieroglyphs), and Greek. The decipherment was largely the work of two people: it started with important insights by the English physicist Thomas Young, then continued and brilliantly concluded by the French linguist Jean-Francois Champollion.

The topic is currently receiving increasing mainstream attention with the development of the semantic web, the relevant areas being ontology alignment, semantic interoperability, open distributed systems, agent communication, distributed reasoning systems, and context-aware systems. Major conferences organized workshops to address these topics [19, 20]. In this context, the current proposal may be relevant to the topics of ontology alignment/mapping, and agent communication. At the same time it is noted that in agents that are purely software the idea of a conceptual system that is grounded in sensory experience seems to be empty. This is discussed in section 4.4.

The author is not aware of a similar formulation or treatment of the problem with mathematical tools as proposed in this article.

2.2 Connections with Other Domains of Research in AI

2.2.1 Machine Learning

The process that is described here could be regarded, in a certain sense, as a form of machine learning [36, 34, 1]. Two agents need to understand each other, but cannot do that, because this necessity could not be anticipated at either one's design time. Machine learning is required for a runtime improvement of either one of the existing designs, to enable this understanding.

In machine learning terminology this could be regarded as a case of learning a function: a 'translating' function. In terms of machine learning, the current framework *biases* this function to be a *p-morphism* as will be defined in the

[1]Bavel, rather than Babel, is the correct transliteration from Hebrew. It means 'confused mix'. The sound 'v' has been distorted into 'b' in the translation of many biblical names through the use of the letter β in an early translation from Hebrew to Greek.

following sections. However, this function that one is looking for is not known at the outset, and the process does not lend itself naturally to a characterization with machine learning terms such as *training set* and *supervised* or *unsupervised* learning. One is rather deploying certain structural constraints that should be satisfied by a yet unknown function, where these constraints are not necessarily presented as values of the function on some training set.

It is perhaps also worthwhile to note that, in spite of that which a Boolean framework might suggest, this is not a study about concept learning. The agents are indeed pointing at samples (examples, instances), and relating them to things that may be regarded as variables (features, attributes, components), or labels (classes, categories, decisions). The use of *true* and *false* may mislead to thinking about positive and negative training instances. However, each one of the participating agents knows its concepts (classifiers, recognizers, categorizers), and this is not about learning new ones or speeding up recognition. The goal is rather to find a mapping between two different *sets of* concepts by using grounding examples and the internal Boolean structures of the two sets of concepts. Disguising this as 'each concept of the other agent being learned as a separate new concept' would be missing the structural benefits in the proposed approach.

2.2.2 Description Logics

The conceptual representation schema that emerges from the formal constructions of ISAAC, as outlined in section 3.2.1, shares structural aspects with Description Logics [16]. A basic shared idea is that intelligent representations and ontologies consist of hierarchies of concepts and relationships among them. This general approach was initially adopted by the first frame systems and semantic networks, then further formalized in the KL-ONE system [29]. Following a seminal paper by Brachman and Levesque [21] it came under the labels 'Terminological Systems', then 'Concept Languages', and finally 'Description Logics'. These stages of development varied in their emphasis on various aspects of representations: epistemological, computational, formal-theoretical, implementational, and so on.

Description Logics follow a long and respectable tradition, that goes back to antiquity, to formalize cognitive processes by capturing rational thought using manipulation of abstract symbols by explicit formal rules. An essence of mathematical modeling and representation is, indeed, about abstractions and formal operations on abstract symbols, which has always been considered, for good reasons, a virtue of these methods. It was taken naturally for granted that the symbols stand for real phenomena that are being abstracted. Then we started delegating the task of manipulating formal representations to non-sensing artificial entities, and the grounding problem surfaced [38, 27].

Due to embodied grounding issues, the very idea of mathematical modeling is often regarded as contradictory to the modeling of representations for highly embodied, situated, agents, being and acting in a real physical world, and innately deriving their intelligence from direct multimodal sensations. AI research has been troubled, not to say plagued, by related controversies for a long time [22, 25, 2, 23, 3].

An underlying idea of ISAAC is to find suitable abstractions that mathematically model precisely that: embodied agents, situated in, and reacting to, an outside environment, and grounding their intelligence in embodied sensations and experiences that are the basis for their representations. Description Logics are hardly concerned with this issue. However, when the embodied grounding issue is properly taken care of, then mathematical methods could still be powerful tools for a formal model of cognitive processes.

Having evolved in parallel with Description Logics, (starting with the preparation of [5]), ISAAC constructs a mathematical structure that starts from a set of sensations/discriminations. That way the representation is modeled to be grounded in the embodiment. Section 3.2.1 outlines how the hierarchies of concepts, and relationships among them, are structured over, and generated by, direct perception. Higher level perception is augmented by Boolean constructs, but everything eventually bottoms out at the basic embodied sensations/discriminations. (The idea that direct experience is the basis of our concepts is argued in [30, 31]).

Please note that this is not at all a claim that cognition invariably works bottom-up. There are indeed significant cognitive processes with a massive top-down component. *It is the underlying conceptual structure that needs to be grounded*, lest these consequential top-down processes might embarrassingly bottom out in meaningless mid-air.

With ISAAC's formal model, different embodiments/implementation platforms are welcome to generate radically different representations. It then becomes of course necessary to grapple with the issue of communication between them, while respecting each one's own grounded meanings. This issue barely comes up in the research about Description Logics.

2.2.3 Boolean Satisfiability

This article proposes to formally model a cognitive process by constructing, and solving, systems of Boolean equations. The solution of systems of Boolean equations is technically related to the propositional satisfiability problem (SAT), which is about deciding whether a given propositional formula is satisfiable. This problem has been recognized as central in quite a few areas such as theoretical computer science, algorithmics, AI, and hardware design and verification. Since the class of satisfiable propositional formulas is NP-complete, there has been a constant quest, with remarkable achievements, for efficient Boolean Satisfiability solvers (An interim overview of that quest is provided in [41]).

In the spirit of mathematical abstraction and context independence that were discussed in the former subsection 2.2.2, research in Boolean Satisfiability is barely concerned with how the relevant Boolean expressions are formed or what they represent. In this article, however, the emphasis is on the grounded construction of the equations themselves, and the integration with other cognitive mechanisms, showing how the solution of the equations further provides a perception morphism, thus modeling a meaningful and rigorous commensuration between two independent perceptions.

A controversy that sometimes comes up in relation to Boolean constructs and SAT in the service of AI, are debates as to the suitability of Boolean operations to model cognition. In their daily life, normal humans do not explicitly construct formulas and solve Boolean equations. Closeness in spirit and apparent behavioral similarity does not necessarily mean similarity in all internal procedural details. The question is whether modeling by a slavish copy of human cognition is required or at all possible. Nobody requires, for instance, neither a slavish copy of biological bodies, nor an exact rehearsal of biological evolution. Instead, one is trying to model by extraction and abstraction of some essence, then letting each implementation platform realize that essence by that which it does best. Artificial platforms do formal calculations better than intuitive short-cuts. The achievement of artificial flight is often mentioned as an example success where copying nature was unnecessary and even unhelpful. (The essence of the connection are the laws of aerodynamics, that each one applies in a different way.)

2.3 Confluence and Integration

ISAAC deploys logic, Boolean algebra, and other mathematical methods, as a scaffold to scale up the conceptual structure from basic sensing-and-reacting embodiments to a higher level conceptual structure, that can support (models of) higher level cognitive processes, including those that are described in section 4 and constitute the core technical contribution of this article.

The proposed methodology could hence be viewed as an integration of a few, some of them seemingly conflicting, mainstream trends in AI, as has been outlined in this contextualization section. The distal goal is a unified theoretical framework for AI, and AI artifact description, analysis, and development within that framework. Rather than developing a new mathematics, the intention is to deploy and to integrate known formal traditions towards a mathematical model, a unified ontology and language of discourse, and systematic AI implementations. Whenever a result is reached, it is examined with regard to pre-theoretical considerations, and tested against existing knowledge and intuitions about the relevant intelligent process. It is invariably asked: Are the results interesting and intuitive for modeling intelligence? If not, then that would be a sign of warning, while meaningful fallout that has not been anticipated at the outset may provide supporting arguments that the model is on a promising track.

3 Specific Context: Basics of ISAAC and its Foregoing Developments

To make the presentation self contained, this section constitutes an abridged, and rather dense, synopsis of ISAAC. The methodology of ISAAC is to have results inferred from formal premises with context-free mathematical rigor. That will be meticulously done in section 4 that provides the core technical contribution of this article. In this outline section, however, there is frequent substitution of detailed mathematical rigor with loose intuitive descriptions. Interested readers are encouraged to further consult the provided references for meticulous detail.

3.1 Formalizing Perceptions

One way of defining and demarcating a mathematical area of discourse is by using the categorical language: The category of sets, with set maps, provides the basis for set theory. The category of groups, with group homomorphisms, provides the basis for group theory, and so on. ISAAC also applies this language, defining a *category of perceptions*, with *perception morphisms* as a basis for the mathematical modeling of perceptions and related cognitive processes. However, no prior knowledge of category theory is required to understand ISAAC. The basic formalism was introduced in [15], along with an extensive discussion of the underlying pre-theoretical intuitions.

Like sets that are the objects of the category of sets, or groups that are the objects of the category of groups, the first definition introduces *perceptions* that are the objects of the basic category that is being constructed. (Explanations follow immediately after the definition.)

Definition 1 *A* Perception *is a three-tuple* $\mathcal{M} = \langle \mathcal{E}, \mathcal{I}, \varrho \rangle$ *where:*

- $\mathcal{E}$ *and* $\mathcal{I}$ *are finite, disjoint sets*
- ϱ *is a 3-valued predicate* $\varrho : \mathcal{E} \times \mathcal{I} \to \{t, f, u\}$[2]

The set $\mathcal{E}$ represents the perceived environment, *world elements/w-elements* that could perhaps be discerned by an agent. Even if the environment exists independent of its perception, then the phenomena that may at all be attended to, and their carving up into individuated w-elements, typically depend on the perceiver: One perceives a forest where another perceives many trees, if at all. Intuitively, the elements of $\mathcal{E}$ can be thought of as indexicals that one uses to relate to things: *this, that...*, and so on. In an AI setting simple examples could perhaps include spatial coordinates, or *'the thing that sensor x currently attends to'*, or *'the input i just received'*, and so on.

The set $\mathcal{I}$ stands for internal impressions/sensations about w-elements. In ISAAC they are called *connotations* or *discriminations*, and they have a subjective existence that is agent specific, enabled and sliced according to the sensory-motor-neural apparatus of the agent. In the general case these are not necessarily overtly expressible, but the context of this paper is about those that can somehow be communicated. Section 4 will use simple examples like 'square', 'rectangle', 'rhombus', and also utterances like 'boo' and 'bla'.

The three-valued predicate ϱ is the *Perception Predicate/p-predicate* that relates between w-elements and discriminations. The intuition behind its values is the following: If $\varrho(w, \alpha) = t$ then w is perceived with discrimination α, if $\varrho(w, \alpha) = f$ then w is perceived to lack discrimination α, and if $\varrho(w, \alpha) = u$ then this perception is undecided about whether w has or lacks discrimination α. The undefined cases might eventually become defined, but right now they are not. [14] provides an extensive discussion about the intuition behind this third truth value. A *total* perception has a *total* p-predicate with no u values.

Actual sets $\mathcal{E}$ and $\mathcal{I}$, and the values of the p-predicate ϱ, once given, model an instantiation of a particular perception. This captures the intuition that perceptions and sensations are innate to agents, determined by their embodiments (as opposed to an external database, for instance). A sensation stands as its own symbol. The definition of $\mathcal{E}$, $\mathcal{I}$, and ϱ avails an open ended diversity of substitution instances for environmental phenomena and their discriminations, including ones for which there are no words and that we may not even know 'what it's like' to perceive and to sense them [35]. That would be determined by each one of the individual instantiations that are being abstracted.

[2] In [15] $\mathcal{E}$ was defined and used, but it was fixed, so $\mathcal{M}$ was defined just by the pair $\langle \mathcal{I}, \varrho \rangle$. In [8, 7] the definition was extended to $\langle \mathcal{E}, \mathcal{I}, \varrho \rangle$, with a variable $\mathcal{E}$.

Given $\mathcal{M} = \langle \mathcal{E}, \mathcal{I}, \varrho \rangle$ as an abstraction of an arbitrary embodied sensory-motor-neural apparatus, the development of the schema proceeds to formalize, bottom-up, what kinds of intelligence could be modeled on top of that, to study the innate capabilities and the suitable modeling tools that would be required for the related constructions, and ways to abstract all that with rigor that can be scrutinized.

Like set maps between sets, or group homomorphisms between groups, commensurations and paths between perceptions are formalized by *perception morphisms (p-morphisms).* (In the categorical language these are often simply called *arrows.*):

Definition 2 *Let* $\mathcal{M}_1 = \langle \mathcal{E}_1, \mathcal{I}_1, \varrho_1 \rangle$ *and* $\mathcal{M}_2 = \langle \mathcal{E}_1, \mathcal{I}_2, \varrho_2 \rangle$ *be perceptions. A* p-morphism $h : \mathcal{M}_1 \to \mathcal{M}_2$ *consists of two set mappings:*

- $h^{\mathcal{E}} : \mathcal{E}_1 \to \mathcal{E}_2$
- $h^{\mathcal{I}} : \mathcal{I}_1 \to \mathcal{I}_2$

These mappings are restricted by the No–Blur *structure preservation condition:*

- $\forall w \in \mathcal{E}$, $\forall \alpha \in \mathcal{I}$, *whenever* $\varrho_1(w, \alpha) \neq u$ *then* $\varrho_2(h^{\mathcal{E}}(w), h^{\mathcal{I}}(\alpha)) = \varrho_1(w, \alpha)$.

A p-morphism is rigid *if the last equality always holds, unconditionally.*

Whenever there is no risk of misunderstanding, the superscripts of $h^{\mathcal{I}}$ and $h^{\mathcal{E}}$ may be omitted.

Composition of p-morphisms and the identity p-morphism are defined on the basis of the composition and the identity of the relevant set mappings. A theorem then follows [15] that perceptions with p-morphisms make a mathematical category, designated $\mathcal{PR}$. This provides a well developed mathematical infrastructure [33, 28, 4] for a 'theory of perceptions'. P-morphisms and categorical constructs with these arrows have been applied to formally model quite a few cognitive processes, as will be outlined below.

3.1.1 Applying Structural Symmetry

One salient property of definitions 1 and 2 is the symmetry between the sets $\mathcal{E}$ and $\mathcal{I}$. From a purely technical, context free, point of view, the roles that a w-element and a discrimination play in the formal definitions are interchangeable. In [8, 7] this symmetry was deployed to parallel, for example, between:

- The mapping $h^{\mathcal{I}} : \mathcal{I}_1 \to \mathcal{I}_2$ of a p-morphism as the *interpretive* component of a transition.
- The mapping $h^{\mathcal{E}} : \mathcal{E}_1 \to \mathcal{E}_2$ of the same p-morphism, as the *literal-analogical* component of the transition.

The first one, $h^{\mathcal{I}}$, we call 'pro-analytic' because it analyzes impressions of cohesive wholes in terms of separate discriminations as basic building blocks, and maps between the discriminations.. The second one, $h^{\mathcal{E}}$, we call 'pro-synthetic' because it takes cohesive w-elements as wholes that are basic building blocks, and maps between these wholes. Formally, however, the schematic construct in both cases is a mapping, and a constructions that is based on this mapping, would look the same in both cases. Hence one could eventually model different cognitive processes with the same formal structure. We call such cognitive processes ‘connatural’, or ‘sibling’, or ’dual’. This *duality* has potential methodical implications:

- Technically, any formal construction or theorem that is established for elements of $\mathcal{I}$ (of $\mathcal{E}$)) can automatically be applied to elements of $\mathcal{E}$ (of $\mathcal{I}$)), mutatis mutandis. (In section 4.7 this will be done for the technical results of this article as well, extending its technical contribution to a novel rigorous modeling of a certain type of creative analogies.)
- In potential computational implementations, the same high-level architectural or computational module, after having been generalized to work with different parameters, could be reused for the ‘sibling’ process: A module that manipulates elements of $\mathcal{I}$ (of $\mathcal{E}$)) could also manipulate elements of $\mathcal{E}$ (of $\mathcal{I}$)), mutatis mutandis. Reusage of modules is a recognized phenomenon in the natural context: Evolution theorists use the term *exaptations* [26] to refer to minor changes that make use of already existing capabilities to create new behaviors. The significance of a capability naturally grows together with the number of behaviors that it supports.
- Taking this even one speculative step further: Since the formalism proposes to model cognitive processes and capabilities, this may suggest looking into the possibility of different cognitive processes being epiphenomenal on the same underlying structure. Such ‘sibling’ cognitive capabilities are expected to emerge side by side. Even if they look superficially unrelated, they may indeed be related in a deeper structural sense. (In section 4.7 this idea will be applied in the context of this article, anticipating that agents that can make out meanings of other agents as described here, could use this capability to make out new environments as well.)

3.1.2 Reactions, Behavior, Affect

The basic definitions 1 and 2 are sufficient for the sequel of this article. It is just noted briefly that, since reactions and behavior typically develop adaptively side by side with perception, there are extensions of this framework to include reactions and behavior (among others: [6, 9, 12]). These extensions are represented by the second A in ISAAC, which stands for ‘affective’. Section 4.6 gives an outline of how the new mechanisms that are contributed by this article can be integrated as a behavior of an autonomous agent.

3.2 Modeling Improved Perception with Boolean Constructs

3.2.1 Basics

Intelligence processes its sensory input using not only ‘its eyes’ (i.e. its sensing equipment), but also ‘its head’, to make sense of its environment, yielding a conceptual system. One of the subsequent steps, following the basic definitions, was hence the modeling of analytic cognitive processes and representations on top of this sense perception framework [14]. To model that, we define, and focus on, a subcategory of the category $\mathcal{PR}$ that was defined above. This subcategory is designated $\mathcal{PR}^{\text{bl-I}}$, and it consists of *Boolean perceptions* of the form $\mathcal{M} = \langle \mathcal{E}, \mathcal{B}, \sigma \rangle$. Boolean perceptions are defined to have sets of discriminations that are Boolean algebras (hence the notation $\mathcal{B}$ instead of $\mathcal{I}$), and their p-predicates, σ, are adequately restricted. For example, the value of $\sigma(w, \alpha \vee \beta)$ is derived from the values of $\sigma(w, \alpha)$ and $\sigma(w, \beta)$ in an expected manner. (As declared at the beginning of this section, this overview does not go into precise definitions, but rather tries to provide an outline of the construction and its guiding intuitions. Full mathematical detail is provided in the references.)

The categorical construction in [14] yields Lukasiewicz-style 3-valued truth tables [32], that may be applied algorithmically for that computation. P-morphisms in the subcategory $\mathcal{PR}^{\text{bl-I}}$ are based on Boolean homomorphisms between the Boolean algebras of discriminations, aligning the representations. A perception with a Boolean algebra of discriminations could use the Boolean structure for the following improvements of its representation and the improvement of cognitive processes that are related to that representation:

- Boolean algebras feature a partial order. This may enable the organization of discriminations in *taxonomic hierarchies*, with inheritance of information.
- The various Boolean operations allow the formation of *compound concepts* as combinations of more basic ones.
- The lattice aspect of Boolean algebras provides links for *ease of access*. (The connection between any two discriminations being through their least upper bound or greatest lower bound.)
- The propositional aspect of Boolean algebras may underlie an interpretation of the representation in logical formulas, and be applied for *ease of inference*.
- The extension of the formalism to reactions and behavior as mentioned in section 3.1.2, regards discriminations *as triggers of reactions*, and Boolean combinations of discriminations then serve purposes of improved integrative behavior (namely an improvement in the control and the management of combined reactions to multiple stimuli). Moreover, in that context it is conjectured that the need to integrate simultaneous stimuli and conflicting reactions in some sensible manner could be an evolutionary pressure behind the development of lattices of compound sensory-motor-neural discriminations, perhaps the precursors of compound concepts and higher level thought.
- *Not the least significant feature of Boolean perceptions is that they allow Boolean equations for making out meanings, and that is the novel technical contribution which constitutes section 4 of this article.*

Following the idea of symmetry from section 3.1.1, a dual Boolean construction was defined with environments: The subcategory $\mathcal{PR}^{\text{bl-E}}$ has perceptions with environments (the $\mathcal{E}$'s) that are Boolean algebras. That enabled the formal modeling and analysis of compound analogies [8] and imaginative design [7]. In a structural sense, they are 'sibling' to analytic and interpretive representational processes. Section 4.7 will discuss the applications of the technical methods from this article for a useful construction of a certain type of creative analogies in the subcategory $\mathcal{PR}^{\text{bl-E}}$.

Finally, in a further restricted subcategory *both environment sets and discrimination sets* are Boolean algebras, modeling perceptions that are capable of both:

- Improved representation and improved cognitive processes that are related to that representation, as detailed in the itemized list above.
- Compound analogies and compound imaginative design.

(These perceptions involve a perception predicate $\mathcal{E} \times \mathcal{B} \to \{t, f, u\}$ that is defined on a product of two Boolean algebras, requiring special caution. The construction revealed an interesting singularity, that can be resolved.)

To avoid tedious notation in the discussions that follow, whenever a 'Boolean construct' is mentioned without further specification, it is intended to be in either one of the subcategories $\mathcal{PR}^{\text{bl-I}}$ or $\mathcal{PR}^{\text{bl-E}}$. The designation $\mathcal{PR}^{\text{bl}}$ should be read as '*either* $\mathcal{PR}^{\text{bl-I}}$ *or* $\mathcal{PR}^{\text{bl-E}}$'.

3.2.2 Validity and Completeness

Notions of validity and completeness in the Boolean constructs model a certain sense of an optimal–under–the–circumstances conceptual system. Among other things, these notions are also necessary for the novel technical constructions in section 4.

Validity and completeness in the context of Boolean perceptions relate between two partial orders:

- The Boolean partial order $\leq$ on constituents, a well known syntactic feature that comes with any Boolean algebra: One writes $\alpha \leq \beta$ whenever $\alpha \wedge \beta = \alpha$ (and in that case also $\alpha \vee \beta = \beta$).
- The perceptual quasi order $\trianglelefteq$ that models observation of lawlike patterns as defined below.

Definition 3 *Let $\mathcal{M} = \langle \mathcal{E}, \mathcal{I}, \varrho \rangle$ be a perception. For all $\alpha, \beta \in \mathcal{I}$ write $\alpha \trianglelefteq \beta$ if, for all $w \in \mathcal{E}$:*

$$\begin{array}{rcl} \varrho(w, \alpha) = t & \Rightarrow & \varrho(w, \beta) = t \\ \varrho(w, \beta) = f & \Rightarrow & \varrho(w, \alpha) = f \end{array} \tag{1}$$

For $w, z \in \mathcal{E}$, $w \trianglelefteq z$ is defined in a dual manner:

Definition 4 *Let $\mathcal{M} = \langle \mathcal{E}, \mathcal{I}, \varrho \rangle$ be a perception. For all $w, z \in \mathcal{E}$ write $w \trianglelefteq z$ if, for all $\alpha \in \mathcal{I}$:*

$$\begin{array}{rcl} \varrho(w, \alpha) = t & \Rightarrow & \varrho(z, \alpha) = t \\ \varrho(z, \alpha) = f & \Rightarrow & \varrho(w, \alpha) = f \end{array} \tag{2}$$

The partial order $\trianglelefteq$ models perceptible lawlike patterns of (Boolean combinations of) constituents. Definitions 3 and 4 are non-monotone with respect to the third truth value u, but without u values, $\trianglelefteq$ coincides with classical 2-valued material implication. A methodological discussion about the choice of that particular definition, which is related to Lukasiewicz's 3-valued logic [32], can be found in [14].

In a valid Boolean perception $\{\leq\} \subseteq \{\trianglelefteq\}$, and in a complete Boolean perception $\{\trianglelefteq\} \subseteq \{\leq\}$. It was shown in [14] that all Boolean perceptions are valid, but not necessarily complete. A typical Boolean perception would be somewhere in-between: some lawlike patterns are observed and internalized, but not all. However, the Boolean subcategories (namely $\mathcal{PR}^{\text{bl-I}}$ and $\mathcal{PR}^{\text{bl-E}}$) do have non trivial subcategories of *valid and complete* perceptions. These perceptions model an idea of acute perception with a thorough observation and total internalization of perceptible lawlike patterns. They model a certain sense of an optimal–under–the–circumstances conceptual systems, that will be relevant for the novel technical constructions in section 4.

3.2.3 Generating Boolean Perceptions

Starting from basic perceptions as in definition 1, an endofunctor of the form $\mathcal{G} : \mathcal{PR} \to \mathcal{PR}$ is applied, where $\mathcal{G}(\mathcal{M})$ is a Boolean perception. That models autonomous cognitive transitions into perceptions that feature higher level capabilities, marrying the grounding provided by the embodied sensory-motor-neural apparatus with the advantages of higher level thought. In [14], two canonical endofunctors of that type were defined and studied. A typical Boolean generation would lie somewhere in between the two. The most general Boolean generation is described first.

3.2.4 Basic Canonical Generation: Free Boolean Perception

The simplest Boolean closure of perceptual constituents takes them to be free generators. The free Boolean generation is defined by a free functor into the Boolean subcategory $\mathcal{G}^{\text{fr}} : \mathcal{PR} \to \mathcal{PR}^{\text{bl}}$. The generating morphism $\xi^{\text{fr}} : \mathcal{M} \to \mathcal{G}^{\text{fr}}(\mathcal{M})$ is a natural transformation from the identity functor on $\mathcal{PR}$ to the functor $\mathcal{G}^{\text{fr}}$. This provides a modeling of a methodical, totally open-minded, general cognitive transition from basic perceptions to Boolean perceptions. However, it has the following drawbacks:

- $\mathcal{G}^{\text{fr}}(\mathcal{M})$ is, in the general case, incomplete. It is impervious to perceptible lawlike patterns of the form $\trianglelefteq$ as in definitions 3 and 4. (Whether a Boolean perception is just valid or also complete is of significance to the novel technical constructions in section 4.)
- If the generating perception $\mathcal{M}$ happens to be Boolean already, then $\mathcal{G}^{\text{fr}}$ unconditionally generates a Boolean set of 2^{2^n} constituents over n generating constituents, blindly duplicating constituents and leading to a combinatorial disaster.

There is a fixed point formalism that serves to characterize these drawbacks. $\mathcal{G}^{\text{fr}}$ does not have a functorial fixed point, whereas the functor that is described next does have that.

3.2.5 Perceptive Canonical Generation: Complete Boolean Perception

Yet another subcategory of $\mathcal{PR}$, designated $\mathcal{PR}^{\text{Sk}}$, is defined to eliminate the source of the perceptual imperviousness of the endofunctor $\mathcal{G}^{\text{fr}}$ from the former section 3.2.4. It enables the modeling of better perceptual acuity. Intuitively, one still wants to do a methodical open-minded general transition to a Boolean closure, but to somehow restrict it to those intelligences that are not totally impervious to the Boolean structure and show a 'natural feel' for that.

The precise formal modeling of that 'natural feel' is captured by the definition of the *sketch structure* that is detailed and studied in [14]. Here we just provide an intuitive description: Consider, for example, a perception $\mathcal{M}$ that features discriminations $\{$*infrared, visible, invisible, ultraviolet*$\}$. It could be impervious to connections between these discriminations, or it could be more acute, observing, for example, that anything is either visible or invisible, that both infrared and ultraviolet are invisible, and so on [3]. More formally: $\neg visible \trianglelefteq invisible$, $infrared \vee ultraviolet \trianglelefteq invisible$, and so on. A sketch structured p-morphism h would then preserve these patterns, so that $\neg h(visible) \trianglelefteq h(invisible)$, and so on.

Because $\mathcal{M}$ from this example is not necessarily Boolean, the only way to observe its acuity, namely its 'natural feel' for Boolean structure, is by looking at the transitions (p-morphisms) that it allows (for instance the way it communicates) rather than in the object perception itself (namely its sensory-motor-neural apparatus).

The subcategory $\mathcal{PR}^{\mathrm{Sk}}$, '*The sketch-structured subcategory of perceptions*', consists of *all* the object perceptions of $\mathcal{PR}$ as in definition 1, but p-morphisms are limited to the more structure preserving type as examplified above: *Sketch Structured p-morphisms* only[4].

Following that, another endofunctor is defined. $\mathcal{G}^{\mathrm{fr-cmp}} : \mathcal{PR}^{\mathrm{Sk}} \to \mathcal{PR}^{\mathrm{bl}}$. It is a free functor that is defined on the sketch-structured subcategory of perceptions, and is shown to yield a complete Boolean perception $\mathcal{G}^{\mathrm{fr-cmp}}(\mathcal{M})$ (whereas $\mathcal{G}^{\mathrm{fr}}(\mathcal{M})$ of section 3.2.4 above was, in the general case, incomplete). With these definitions one gets the further restricted Boolean subcategory $\mathcal{G}^{\mathrm{fr-cmp}}(\mathcal{PR}^{\mathrm{Sk}}) = \mathcal{PR}^{\mathrm{bl-cmp}}$.

To summarize loosely, the endofunctor $\mathcal{G}^{\mathrm{fr-cmp}}$ is based on a construction that 'moves things around' in the generated Boolean closure so that perceived patterns should be reflected in this generated Boolean closure. Consequently, the perceptions in the subcategory $\mathcal{PR}^{\mathrm{bl-cmp}}$, are *valid and complete* Boolean perceptions that may be generated over all basic perceptions $\mathcal{G}^{\mathrm{fr-cmp}}(\mathcal{M})$.

The sketch structured generating p-morphism $\xi^{\mathrm{Sk}} : \mathcal{M} \to \mathcal{G}^{\mathrm{fr-cmp}}(\mathcal{M})$ is a natural transformation from the identity functor on $\mathcal{PR}^{\mathrm{Sk}}$ to the functor $\mathcal{G}^{\mathrm{fr-cmp}}$. This provides the modeling of a methodical cognitive transition from basic perceptions to Boolean perceptions that is still fairly open-minded and general[5], but this one is also *perceptually acute* because:

- $\mathcal{G}^{\mathrm{fr-cmp}}(\mathcal{M})$ is valid and complete: there is a thorough observation and total internalization of perceptible lawlike patterns.

- For all valid and complete Boolean perceptions $\mathcal{M}$, $\mathcal{G}^{\mathrm{fr-cmp}}(\mathcal{M}) = \mathcal{M}$. Namely: If the generating perception happens to already be a valid and complete Boolean perception, then $\mathcal{G}^{\mathrm{fr-cmp}}$ 'behaves as if it knows that', modeling a certain notion of self awareness.

[3] A similar approach to the intelligent analysis of data was independently developed in [18].

[4] a subcategory with the same objects, but perhaps fewer morphisms, is sometimes called a *wide* or a *lluf* subcategory [40, p.211].

[5] The no-free-lunch price is non-monotonicity: Indeed, if some perceptible lawlike pattern breaks, one has to trace back to the drawing board and regenerate a modified Boolean structure.

There is a functorial fixed point formalism that serves to characterize these advantages, where valid and complete Boolean perceptions are characterized as functorial fixed points of $\mathcal{G}^{\mathrm{fr-cmp}}$.

As mentioned before, whether a Boolean perception is just valid or also complete is of significance to the core technical contribution in the next section.

4 Unravelling Meanings with Boolean Equations

4.1 The Problem

Section 3 outlined (on the basis of the detailed articles that it referenced) how every perception has its own perceptual discrimination system, and p-morphisms translate between different perceptions. Meanings are preserved across p-morphisms by the structure preservation condition: Transitions between discriminations are grounded by commensurate w-elements, and, on the other hand, analogies between w-elements need to be justified by commensurate discriminations. The question that we now ask is how to obtain a translating p-morphism when none is provided. We start with a simple working example from human interaction, describing how humans intuitively overcome semantic heterogeneity in linguistic communication, without having any privileged access 'into each other's head'.

4.2 A Working Example

A traveler is on a train in a foreign country whose language she does not speak, passing the time by playing with a child who folds and refolds the train tickets, fabricating various geometrical forms. The traveler calls these forms: 'square', 'rectangle', 'rhombus'. The child utters 'boo', 'bla', and other unintelligible tokens. After a while the traveler comes to the conclusion that maybe 'bla' means 'right angles', and 'boo' means 'all sides equal'. She proceeds to fold the tickets into more forms, to test the conjecture and to find the meaning of more tokens.

The conjecture is not merely reached by pure inspiration. During the interaction, they are consistently pointing at things and calling them names. They also use universal face expressions that mean 'yes', 'no', or hesitation [24]. When one collects a few shapes together, the other assumes that the naming that follows applies to the collection, and so on. Is it possible to formulate a domain general methodology, and to determine the required infrastructure, so that it could be rigorously applied for unravelling meanings in an arbitrary situation between two arbitrary perceptions?

Although the two participants in the example story do not share a language in a narrow sense, it is clear that they do share something. Our intuition says that the shared core has to do with 'yes', 'no', constructing collections, and so on. These are roughly the things that George Boole called *The Laws of Thought* when he undertook their formal investigation [17]. It is proposed that what the traveler was doing in that train can be modeled by solving a system of Boolean equations.

4.3 Solving the Example

Let creative communication yield that a rhombus is 'boo' (and other unintelligible things), a rectangles is 'bla' (and other unintelligible things), and a square is exactly 'boo' and 'bla'. One gets the following system of Boolean equations:

$$\begin{array}{rcl} \textit{rhombus} & \leq & \textit{boo} \\ \textit{rectangle} & \leq & \textit{bla} \\ \textit{square} & = & \textit{boo} \wedge \textit{bla} \end{array} \tag{3}$$

A detailed study of Boolean equations can be found in [37]. It analyzes how Boolean equations can be solved, algorithmically. Complexity issues are also dealt with. Efficient SAT techniques can be applied here as well. The solution of a system of Boolean equations is not always unique, and in that case it is possible to arrive at all possible solutions. Quite a few examples are analyzed in [37], including the following solution of (an equivalent of) the system (3):

$$\begin{array}{rcl} \textit{boo} & = & \textit{rhombus} \vee \textit{square} \\ \textit{bla} & = & \textit{square} \vee \textit{rectangle} \end{array} \tag{4}$$

The system (3) of Boolean equations happens inside a Boolean subalgebra of discriminations from the traveler's perception, where *square, rectangle, rhombus* are constant discriminations, and *boo, bla* are unknowns. The solution (4) provides discriminations from the traveler's perception as values for these unknowns.

Finally, the traveler's Boolean perception features the following equalities inside its own Boolean algebra of discriminations:

$$\begin{array}{rcl} \textit{rhombus} \vee \textit{square} & = & \textit{all\_sides\_equal} \\ \textit{square} \vee \textit{rectangle} & = & \textit{right\_angles} \end{array} \tag{5}$$

This last step is an inference inside the traveler's own conceptual system.

4.4 Methodological Assumptions

The following methodological assumptions are implied in the descriptions of the working example and its solution (sections 4.2 and 4.3 above). They will also be assumed for the general treatment.

- We assume that the perceived environment under consideration (e.g. the folded train tickets) is shared by the two sides.
- It is also assumed that as long as the described process is going on, there will be no state changes (namely changes in the environment or in the perceptions of the participants)) that might effect the process and its solution.
- It is assumed that each one of the participants is able to observe certain relevant activities of the other participant. In particular there is a possibility of 'pointing to' things. This further implies the following methodological assumptions:

– The process is intended for agents with a certain embodiment and some minimal sensory-motor-neural capabilities, and not for software agents. As a matter of fact, in agents that are purely software the idea of a conceptual system that emerges from, and is grounded in, own sensory-motor-neural experience seems to be empty, and hence not directly relevant to the present context.

– The process is not intended for starting out with abstract concepts, at least not directly. Abstract concepts are typically generalizations of collections of concrete concepts. In the Boolean lattice framework of ISAAC they would often be modeled by a least upper bound. For example, the number 3 would be modeled by a disjunction of '3 apples or 3 oranges or ... ' and so on. In the working example, the traveler arrives at the generalization in equation (5) only at the last stage, applying an inference in its own conceptual system.

  Indeed, the traveler would be ill advised to directly head for the meaning of concepts like 'democracy', or 'freedom', or 'beauty', but would rather start with the understanding of concrete examples that might eventually generalize into these concepts.

Further methodological issues will naturally emerge as we proceed, and they will hence be duely considered towards the end, in the discussion section 4.8.

4.5 The General Solution

4.5.1 Getting Started

One needs to model observations like '*a rhombus is boo (and perhaps more unintelligible things)*' as Boolean equations (in a Boolean algebra of discriminations of a perception).

We start with the following technical requirements (that will be relaxed later):

- At least one of the agents is an *active agent*. (In the working example this is instantiated by the traveler.) An active agent should have a complete Boolean perception, as defined in section 3.2.2, that is designated $\mathcal{M}_0 = \langle \mathcal{E}, \mathcal{B}_0, \sigma_0 \rangle$. This is required for definition 6, as explained in remark 5 that follows that definition. This requirement to have a complete Boolean perception will be relaxed later, in section 4.8.2.

- The active agent is trying to figure out the meanings of the other agent. (In the working example the other agent is instantiated by the local child.) This other agent should have a valid Boolean perception, as defined in section 3.2.2, that is designated $\mathcal{M}_2 = \langle \mathcal{E}, \mathcal{B}_2, \sigma_2 \rangle$. This requirement to have a Boolean perception will also be relaxed later, in section 4.8.2.

To initialize the process, one needs to model the fact that the active agent comes at all to perceive how the other agent discriminates. In the working example, the traveler comes to perceive that the local child utters 'boo', 'bla', and so on. That is modeled by first transiting from $\mathcal{M}_0$ to $\mathcal{M}_1$ by adding discriminations of the form: '*this is* β *for* $\mathcal{M}_2$ ', abbreviated as $\overline{\beta}$. Namely: discriminations with an overline stand for observed discriminations from $\mathcal{M}_2$. (In the working example one wants to add $\overline{boo}$ and $\overline{bla}$ to the traveler's perception.) In the categorical framework, changes and transitions are formalized by arrows, namely p-morphisms. In this case $g : \mathcal{M}_0 \to \mathcal{M}_1$ will be composed of two arrowed steps.

1. The first arrow simply adds the overlined discriminations to $\mathcal{M}_0$, namely: $\mathcal{M}_0 \to \langle \mathcal{E}, \mathcal{B}_0 \cup \{\overline{\beta}\}, \varrho \rangle$ such that:

   - The mapping of discriminations simply embeds $\mathcal{B}_0$ in the union: $\mathcal{B}_0 \hookrightarrow \mathcal{B}_0 \cup \{\overline{\beta}\}$.
   - For discriminations in $\mathcal{B}_0$, the p-predicate ϱ is the same as σ_0.
   - For the added discriminations, namely $\{\overline{\beta}\}$, the p-predicate is naturally defined as follows: $\varrho(w, \overline{\beta}) = t$ if $\mathcal{M}_2$ seems to react $\overline{\beta}$ to w, f if not, and u if this is not clear.
   - The transition $h^{\mathcal{E}}$ is the identity on $\mathcal{E}$.

2. The second arrow is the natural transformation $\langle \mathcal{E}, \mathcal{B}_0 \cup \{\overline{\beta}\}, \varrho \rangle \to \mathcal{M}_1$, that generates the complete Boolean perception $\mathcal{M}_1 = \langle \mathcal{E}, \mathcal{B}_1, \sigma_1 \rangle$ over the codomain of the first arrow (see section 3.2.5).

It is easy to see that the difference between $\mathcal{M}_0$ and $\mathcal{M}_1$ bottoms out exactly at the added discriminations of the form $\overline{\beta}$, and that the composite arrow $g : \mathcal{M}_0 \to \mathcal{M}_1$ is a Boolean p-morphism (it simply embeds $\mathcal{M}_0$ in $\mathcal{M}_1$).

Remark 1 *The discriminations* 'this is β for $\mathcal{M}_2$ '*, abbreviated as $\overline{\beta}$, are amodal: β could stand for any observed reaction, including nonverbal ones such as various gestures.*

Remark 2 *A unified account of autonomous agents should also model how processes are triggered, namely, in the present context, how the transition $\mathcal{M}_0 \to \mathcal{M}_1$ is initiated, as well as the processes that follow below. This will be done in section 4.6.*

Given $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$ as above, the next step is to formalize the active agent's observations as lawlike patterns in $\mathcal{M}_1$ (see definition 3 in section 3.2.2). In the working example these would be the following lawlike patterns in the traveler's perception:

$$\begin{array}{rcl} rhombus & \trianglelefteq & \overline{boo} \\ rectangle & \trianglelefteq & \overline{bla} \\ square & = & \overline{boo} \wedge \overline{bla} \end{array} \tag{6}$$

For example, the first one holds, by definition 3, because for all $w \in \mathcal{E}$:

$$\begin{array}{rcl} \sigma_1(w, rhombus) = t & \Rightarrow & \sigma_1(w, \overline{boo}) = t \\ \sigma_1(w, \overline{boo}) = f & \Rightarrow & \sigma_1(w, rhombus) = f \end{array} \tag{7}$$

Remark 3 *One can see now that the transition $\mathcal{M}_0 \to \mathcal{M}_1$ was needed to provide a perception within which the lawlike patterns are at all observed.*

Remark 4 *The treatment of this section works with all truth values, including lawlike patterns that hold just by virtue of: $\sigma_1(w, \overline{\beta}) = f$ or $\sigma_1(w, \overline{\beta}) = u$. Therefore the discussion will not be restricted otherwise. In practice, however, it does not seem sensible to make an observation just by virtue of things being undefined or false, since at the end of the day that will add to the list of solutions too many 'theoretically possible' options that are not really justified. This issue will be discussed again in the fallout section 4.8.*

In a typical case, the discriminations on one side of observations such as (6) would be without overlines (e.g. the left hand sides in (6)) while the other side would consist of overlined discriminations (e.g. the right hand sides in (6)). However, that restriction is not really necessary, as the equations are all operated inside $\mathcal{B}_1$, so one is allowed to formulate 'mixed' lawlike patterns.

Definition 5 *Let $\mathcal{M}_1 = \langle \mathcal{E}, \mathcal{B}_1, \sigma_1 \rangle$ and $\mathcal{M}_2 = \langle \mathcal{E}, \mathcal{B}_2, \sigma_2 \rangle$ be the Boolean perceptions as described above. A* 'Rosetta' *of $\mathcal{M}_1$ for $\mathcal{M}_2$ consists of $n \geq 1$ lawlike patterns of the form:*

$$f_\imath \ \star_\imath g_\imath \tag{8}$$

Where:

- $\star_\imath \in \{ = , \trianglelefteq , \trianglerighteq \}$.
- $\{f_\imath\}_{\imath=1,n}$ *and* $\{g_\imath\}_{\imath=1,n}$ *are Boolean expressions in* $\mathcal{B}_1$.

From the observations of a rosetta one proceeds to obtain a system of Boolean equations:

Definition 6 *Let $\mathcal{M}_0$, $\mathcal{M}_1$, and $\mathcal{M}_2$ be Boolean perceptions as above. Let there be a rosetta of $\mathcal{M}_1$ for $\mathcal{M}_2$ as above. The corresponding* 'system of equations' *is a system of n simultaneous Boolean equations in $\mathcal{B}_0$, which is obtained as follows:*

- *Define φ by*

$$= \overset{\varphi}{\mapsto} = \quad , \quad \trianglelefteq \overset{\varphi}{\mapsto} \leq \quad , \quad \trianglerighteq \overset{\varphi}{\mapsto} \geq \tag{9}$$

- *The $\overline{\beta}$'s are taken as unknown variables in $\mathcal{B}_0$.*
- *The i'th Boolean equation is then written as: $f_\imath \ \varphi(\star_\imath) \ g_\imath$.*

For example, the rosetta (6) yields (3) as its system of Boolean equations.

Remark 5 *Definition 6 requires that $\mathcal{M}_0$ and $\mathcal{M}_1$ be* complete *Boolean perceptions, so that observed patterns can be replaced by the Boolean partial order as in (9). This is the reason behind the first technical requirement that is itemized at the beginning of section 4.5. This requirement will be relaxed in section 4.8.2.*

4.5.2 Getting Through

The following theorem is the backbone of the modeled cognitive process:

Theorem 1 *Let there be a rosetta of $\mathcal{M}_1$ for $\mathcal{M}_2$ as in definition 5. If the corresponding system of equations has a solution where every unknown $\overline{\beta}$ of the system can be replaced by h_β that is a constant expression in $\mathcal{B}_0$, such that the system is satisfied, then for every such solution:*

1. *The mapping $h : \beta \mapsto h_\beta$ can be extended to a Boolean homomorphism $h : \mathcal{B}_2' \to \mathcal{B}_0'$, where $\mathcal{B}_2' \subseteq \mathcal{B}_2$ is the subalgebra generated by the β's of the rosetta, and $\mathcal{B}_0' \subseteq \mathcal{B}_0$ is the subalgebra generated by the h_β's.*

2. *There exists a fully specified perception* $\widehat{\mathcal{M}'_2} = \langle \mathcal{E}, \mathcal{B}'_2, \widehat{\sigma_2} \rangle$ *that, for all one knows, can be substituted for* $\mathcal{M}_2$ *without causing any change in the entire construction from the beginning of this section 4.5, and the homomorphism* h *defines a rigid p-morphism*[6] $h : \widehat{\mathcal{M}'_2} \to \mathcal{M}'_1$.

This theorem (its proof follows immediately below) actually tells us exactly how solving a system of equations provides:

- A perception $\widehat{\mathcal{M}'_2}$ that is a conjectured approximation for $\mathcal{M}'_2$ (in the working example: a conjectured approximation of the perception of the local child), and also:
- A corresponding conjectured translation h (in the working example: a conjectured translation from the local child's words to the traveler's words).

Proof.

1. Item 1 of the theorem is proven first. Boolean algebra tells us[7] that h could be extended to a Boolean homomorphism if and only if the following holds: For every sequence $\beta_1, \beta_2, ...\beta_k$ from the β's, and for every sequence $\eta_1, \eta_2, ...\eta_k$ of numbers $-1, 1$:[8]

$$\eta_1\beta_1 \wedge ... \wedge \eta_k\beta_k = \bot_{\mathcal{B}'_2} \tag{10}$$

implies that:

$$\eta_1 h(\beta_1) \wedge ... \wedge \eta_k h(\beta_k) = \bot_{\mathcal{B}'_1} \tag{11}$$

To show that, first note that, by the validity of the Boolean perception $\mathcal{M}_2$, (10) above implies that: $\forall w \in \mathcal{E}\ \sigma_2(w, \eta_1\beta_1 \wedge ... \wedge \eta_k\beta_k) = f$. Hence, by the validity of the Boolean perception $\mathcal{M}_1$, and without loss of generality, the following can be added to the rosetta, and to the system of equations: $\bot_{\mathcal{B}'_1} = \eta_1\overline{\beta_1} \wedge ... \wedge \eta_k\overline{\beta_k}$. Since the unknowns of this equation can be replaced by their solutions, that implies (11), and that completes the proof of item 1 of the theorem.

2. To prove item 2 of the theorem: Based on 1, then for all α in $\mathcal{B}'_2$, one can consistently add to the system of equations, without loss of generality, all the equations of the form: $h(\alpha) = \overline{\alpha}$. That completes the proof of item 2 of the theorem.

□

As already said, this result provides the perception $\widehat{\mathcal{M}'_2}$ as a conjectured approximation, for all one knows, for $\mathcal{M}'_2$, and h as the corresponding conjectured translation. Moreover, the 'observations' of the form $h(\alpha) = \overline{\alpha}$ are *testable predictions* that could be applied to test the conjecture. It is easy to see that the conjecture would still hold, with the suitable slight blurring, if for some $w \in \mathcal{E}$ and some $\alpha \in \mathcal{B}'_2$, $\sigma_2(w, \alpha) = u$ but $\widehat{\sigma_2}(w, \alpha) \neq u$[9]. In addition, a rigid p-morphism h has an inverse h' such that $h' \circ h$ is the identity on $h(\mathcal{B}'_1) \subseteq \mathcal{B}'_2$, so this 'translation' can be tested both ways.

[6] Rigid p-morphisms are defined in the last part of definition 2.

[7] See [39, p.36].

[8] Following [39] we use the notation $(-1)\alpha = \neg\alpha$, and $(+1)\alpha = \alpha$.

[9] If one takes a minute to think about that: The local child might be surprised to learn something about its own conceptual system from a foreigner...

The following theorem is an inverse of theorem 1. It tells us that whenever there exists a translation, then one should be able to figure it out using the described formal process. Namely: there is a collection of observations and corresponding equations, that, when solved, would yield that translation.

Theorem 2 *If there exists a rigid Boolean p-morphism $h : \mathcal{M}_2 \to \mathcal{M}_1$, then there exists a rosetta for $\mathcal{M}_1$ and $\mathcal{M}_2$, that yields h as above.*

Proof. The rosetta would consist of the observations $\forall w \in \mathcal{E} \quad \sigma_2(w, \alpha) = \sigma_1(w, h(\alpha))$ for all (or just a set of generators) α in $\mathcal{B}'_2$. $\square$

An example: The universal perception of $\mathcal{E}$, $\mathcal{U}_{\mathcal{E}} = \langle \mathcal{E}, 2^{\mathcal{E}}, \epsilon \rangle$, is a lax terminal object of the subcategory of the perceptions of $\mathcal{E}$ (see [15]). $\mathcal{U}_{\mathcal{E}}$ is a complete Boolean perception, with a totally two valued p-predicate, and a unique discrimination to describe every subset in the environment. (Its terminal property is lax because arrows to $\mathcal{U}_{\mathcal{E}}$ are not unique in the 3-valued context.) If arrows $\mathcal{M} \to \mathcal{U}_{\mathcal{E}}$ are restricted to their rigid part, then a translation, and a rosetta, from the relevant subperception of $\mathcal{M}$ into the universal perception $\mathcal{U}_{\mathcal{E}}$ always exists.

4.6 Integrating the Process as an Autonomous Behavior

The full framework of ISAAC also proposes to model how processes are autonomously triggered, most typically as reactions to certain discriminations. In this case the novel processes from this section should be autonomously triggered whenever: ($\imath$) An agent observes (either accidentally or intentionally) another agent that produces unintelligible signals, and ($\imath\imath$) The observing agent is motivated or curious enough to try and understand the meanings of these signals, and ($\imath\imath\imath$) The observing agent is also capable enough, namely it is able to perform the procedures that are described in this section.

Below is an outline of how the technical contribution of this article can be integrated into the unified schema of ISAAC.

- Things would start with the (often unsolicited) appearance of the relevant novel environmental phenomena, namely unfamiliar actions of another agent (e.g. someone who says ‘boo’ and ‘bla’).

- New phenomena that enter into the environment are invariably modeled by a p-morphism that extends the environment with the new phenomena, simply embedding the domain environment into an extended one that features w-elements like the indexical *what_he_just_said*.

- The unrecognized novel phenomena would be perceived with a discrimination such as *i_don’t_understand*. It could be abbreviated as a dedicated discrimination *???*.

- In the extended formalism discriminations trigger reactions. Reactions could be of various natures, and not only physical or overt. Internal ‘mental’ reactions could consist, in particular, of intra-agent p-morphisms. In ‘curious’ agents, the discrimination *???* would trigger a cognitive process of ‘curiosity’ that tries to make sense and sort out the phenomena that are perceived with the discrimination *???*, by initiating the transition $h : \mathcal{M}_0 \to \mathcal{M}_1$ that is described at the beginning of section 4.5. (Of course, agents are not necessarily curious, and curiosity could manifest itself in other ways as well. It all depends on how the agent is structured and programmed.)

- The detection of overlined discriminations such as the $\overline{\beta}$'s in $\mathcal{M}_1$, that are 'unknown', could trigger 'curious reasoning', namely the chain of processes that is described in section 4.5:
- An internal process that looks, inside one's own perception $\mathcal{M}_1$, for lawlike patterns that involve the $\overline{\beta}$'s.
- That would generate a rosetta.
- That would generate the equations.
- That would initiate a process that finds the solution of the equations.
- That would finally provide the conjectured approximation $\widehat{\mathcal{M}'_2}$ for $\mathcal{M}'_2$ and the corresponding conjectured translation h.
- If there is more than one solution, that would trigger a choice as suggested in section 4.8.1.

Note that these steps lead to one agent conceiving of the perception of another agent. The possibility of nesting perceptions is theoretically challenging. It opens the way for the modeling of various aspects not only of social cognition, but also of meta-cognition, with the nested construct being recursively one's own. [10, 11, 13] grapple with that option within ISAAC.

4.7 Additional Result by Symmetry: Trying to Understand an Unknown Environment

In section 3.1.1 a *duality* property was observed, based on the fact that the roles that a w-element and a discrimination play in the definitions are technically interchangeable, so that any formal result that is established for $\mathcal{I}$ can automatically be applied to $\mathcal{E}$, mutatis mutandis. One then obtains a formal 'sibling' process, that would possibly model a cognitive process that is quite different from the original process that gave rise to the construction. We now apply this principle to extend the scope of the novel contribution of this paper to model another cognitive process, that is dual in this structural sense.

The core technical contribution of this paper was about obtaining a p-morphism that consists of an interpretive translation of the form $h^{\mathcal{I}} : \mathcal{I}_1 \to \mathcal{I}_2$ where $h^{\mathcal{E}}$ is the identity, namely the two perceptions share the same environment $\mathcal{E}$. If that is formally repeated for the 'sibling' process, one obtains a literal analogy of the form $h^{\mathcal{E}} : \mathcal{E}_1 \to \mathcal{E}_2$, for two perceptions that share the same discriminations $\mathcal{I}$. This would model obtaining an analogy between an unknown environment and a known one, while anchoring the (possibly metaphorical) discriminations that need to be preserved by the analogy. The need for that can indeed arise when an agent has to figure out an unknown environment by finding an analogy to a familiar environment. The transition would then be intra-agent, rather then inter-agent, with equations that involve w-elements rather than discriminations.

A fallout is a prediction that an intelligence that is technically adept at making out meanings for purposes of communication with other agents, as described in this paper, would also be able to adapt to new environments by constructing analogies with more familiar environments. The same technical capability is required for both skills. All one needs to do is to abstract the functionality of the module that implements the first capability, and apply it for the second capability.

The initiation and activation of the analogizing process would be dual to the autonomous process that is described in 4.6 above.

Indeed, translations and analogies have a lot in common, and ISAAC does capture that because they are sibling processes in this formalism. (ISAAC's formalization of analogies and metaphors can be found in [8].)

4.8 Discussion

4.8.1 Sorting the Solutions

What if there is more than one solution to the rosetta system? Following remark 4, this may be, in part, the result of observations that are valid by virtue of too many u or f values, so one might be able to reduce the number of solutions by restricting the rosetta to 'definite' observations that involve mostly t values, less f values, even less u values, and absolutely no observations that are supported only by things being undefined or false.

Solutions may be adopted or discarded by further observations that augment the system of equations. Theorem 1 provides testable predictions that could be applied to test a conjectured solution.

Complex agents that perceive complex environments may still feature too many discriminations, with too many equations and too many solutions to handle all at once. Even for intelligent human beings, making out a new culture and a new language is far from being an instant process, to the extent that if this proposal had come up with an invariably fast process, that would have raised serious doubts about the cognitive plausibility of the proposed approach. It could be a good idea to start with small sub-perceptions, perceiving sub-environments with restricted sets of discriminations, as done in the working example. Then the smaller sub-perceptions may eventually be integrated into a broader perception, providing a bigger picture. In ISAAC's categorical formalism such an integration is readily modeled by a p-pushout, as shown in [15].

Lastly, recall *Ockham's Razor* principle, often cited by researchers in AI: the most likely hypothesis is the simplest one that is consistent with all observations[10], for example a solution that involves the least complex Boolean expressions.

4.8.2 Relaxing Requirements

A look back at the construction (specifically at definition 6 and remark 5 that follows that definition), shows that making out meanings as described above requires that the perception of the active agent ($\mathcal{M}_0$ that transits into $\mathcal{M}_1$), within which the system of equations is solved, should be a complete Boolean

[10] 'Entities are not to be multiplied beyond necessity': *entia non sunt multiplicanda praeter necessitatem*, William of Ockham (or Occam), 1285 – 1349, English philosopher.

perception. (See also the technical requirements at the beginning of section 4.5.) On one hand, this adds to the collection of simple intuitions that are rigorously systematized by ISAAC's formalism: The onus of understanding is on the more intelligent partner. At the same time, some requirements may be slightly relaxed:

- $\mathcal{M}_0$ and $\mathcal{M}_1$ are required to be *complete* Boolean perceptions, which does not sound realistic. (As already remarked in section 3.2.2, in a typical Boolean perception some lawlike patterns are observed and internalized, but not all.) However, by following the details of definition 6, this requirement may be relaxed: only the lawlike patterns that are observed in the rosetta are replaced by the Boolean partial order, so only they need to be internalized in the Boolean structures of $\mathcal{M}_0$ and $\mathcal{M}_1$.
- The other perception, namely $\mathcal{M}_2$, is required to be a Boolean perception. However, by following the construction, and the proof of theorem 1, it can be observed that it is actually the active agent that conceives of the other perception $\mathcal{M}_2$ as Boolean, in order to do its job, but $\mathcal{M}_2$ does not have to be Boolean in itself.

4.8.3 Issues for Further Investigation

- This study is theoretical, describing the mathematics of a proposed method. A next step would be to apply the method to a realistic testbed. One would then be able to study how the approach scales to larger examples. Among others, issues for study would then be the applicability of SAT methods to speed up the solution of equations, and whether the proposed heuristics from section 4.8.1 are sufficient for the proposal to yield a reasonably restricted number of solutions.
- A question that may be looked into is about cases where both perceptions happen to be upscaled enough, namely to feature all the formal requirements, to be the active partner and to do the job. In that case, can one get a significant improvement?
- Another possible direction of further research is about making observations. In the working example, intuitive human interaction produced the observations that defined the rosetta. It is possible to ask questions at random to produce observations from scratch, but is there a methodical way to go about this free exploration that is better in some sense?

5 Conclusion

The problem that is addressed in this article is about obtaining a translation between two different conceptualizations, for purposes of intelligent communication, where no prior anticipation or provisions have been made to prepare for communication and translation between them.

The proposed theoretical solution to the problem is original, and operated within the general framework of ISAAC (*I*ntegrated *S*chema for *A*utonomous *A*ffective *C*ognition), which is a methodologically oriented, long term research that uses algebraic and categorical formalisms with the goal of setting the modeling of autonomous intelligent agents on a unified and rigorous mathematical basis. With this formalism the problem is cast in a rigorous form of a well defined mathematical question: Given two objects a given category, one is looking to define a morphism between them.

A working example is provided where the problem is solved by two human agents with different conceptual systems. This provides pre-theoretical intuitions for the rigorous formal process that is the core novel contribution of this article.

The proposed theoretical solution process is domain general, and it consists of constructing, and solving, Boolean equations that are grounded in the shared environment of the two agents. The process yields a testable conjectured perception of the other, along with a testable conjectured translation that is a perception morphism from that other perception to one's own. Theorems are proven that show the mathematical validity of the construction.

It is also shown that the new formal construction can also be applied to formalize another cognitive process: Obtaining an analogy between an unknown environment and a known one for purposes of intelligent adaptation to a new environment.

It is also shown how the two newly formalized processes can be embedded in the unified schema of ISAAC, and autonomously triggered when necessary.

The presentation of the technical contributions of the article wraps up with a discussion about relaxing some technical requirements that agents need for applying the process, and suggested heuristics to handle cases of multiple solutions to the system of Boolean equations. Issues and directions for further research are discusses as well.

References

[1] E. Alpaydin. *Introduction to Machine Learning*. MIT Press, 2004.

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

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

[4] M. A. Arbib and E. G. Manes. *Arrows, Structures and Functors – The Categorical Imperative*. Academic Press Inc, New York, 1975.

[5] Z. Arzi-Gonczarowski. *Categorical Tools for Artificial Perception*. PhD thesis, Hebrew University of Jerusalem, 1994.

[6] Z. Arzi-Gonczarowski. Wisely non rational – a categorical view of emotional cognitive artificial perceptions. In D. Cañamero, editor, *Emotional and Intelligent: The Tangled Knot of Cognition*, pages 7–12. AAAI Press Technical Report FS-98-03, 1998.

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

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

[9] Z. Arzi-Gonczarowski. A categorization of autonomous action tendencies: The mathematics of emotions. In *Cybernetics and Systems 2000*, volume 2, pages 683–688. Austrian Society for Cybernetic Studies, Vienna, 2000.

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

[11] Z. Arzi-Gonczarowski. Self, empathy, manipulativity: Mathematical connections between higher order perception, emotions, and social cognition. In D. Cañamero, editor, *Emotional and Intelligent II – The Tangled Knot of Social Cognition*, pages 9–14. AAAI Press Technical Report FS-01-02, 2001.

[12] Z. Arzi-Gonczarowski. AI emotions: Will one know them when one sees them? In *Cybernetics and Systems 2002*, volume 2, pages 739–744. Austrian Society for Cybernetic Studies, Vienna, 2002.

[13] Z. Arzi-Gonczarowski. Metacognition the mathematical way: Trying to nest constructs. In M. Anderson and T. Oates, editors, *Metacognition in Computation*, pages 24–29. AAAI Press Technical Report SS-05-04, 2005.

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

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

[16] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation and Application*. Cambridge University Press, 2003.

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

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

[19] P. Bouquet, editor. *Meaning Negotiation: Papers from the* AAAI *Workshop*. AAAI Press Technical Report WS-02-09, 2002.

[20] P. Bouquet and L. Serafini. ISWC-04 workshop on meaning coordination and negotiation. http://dit.unitn.it/ bouquet/ISWC-04-MCN/, November 2004.

[21] R. J. Brachman and H. J. Levesque. The tractability of subsumption in frame-based description languages. In R. J. Brachman, editor, *Proceedings of the 4th National Conference on Artificial Intelligence*, page 34, Austin, TX, August 1984. AAAI Press.

[22] R. A. Brooks. Intelligence without representation. *Artificial Intelligence, special volume on Foundations of AI*, 47:139–159, 1991.

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

[24] C. Darwin. *The Expression of the Emotions in Man and Animals*. Oxford University Press, 1998.

[25] K. M. Ford, P. J. Hayes, and N. Agnew. On babies and bathwater: A cautionary tale. AI *Magazine*, 15(4):14–26, 1994.

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

[27] S. Harnad. The symbol grounding problem. *Physica*, D 42:335–346, 1990.

[28] H. Herrlich and G. E. Strecker. *Category Theory*. Allyn and Bacon, 1973.

[29] R. J.Brachman and J. G. Schmolze. An overview of the KL-ONE knowledge representation system. *Cognitive Science*, 9(2):171–216, 1985.

[30] G. Lakoff and M. Johnson. *Metaphors We Live By*. The University of Chicago Press, 1980.

[31] G. Lakoff and M. Johnson. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books, New York, 1999.

[32] J. Lukasiewicz. On 3-valued logic (1920). In S. McCall, editor, *Polish Logic*. Oxford University Press, 1967.

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

[34] T. Mitchell. *Machine Learning*. McGraw Hill, 1997.

[35] T. Nagel. What is it like to be a bat? In N. Block, editor, *Readings in Philosophy of Psychology*. Harvard University Press, 1980.

[36] N. J. Nilsson. Introduction to machine learning. http://ai.stanford.edu/ nilsson/mlbook.html, 1996.

[37] S. Rudeanu. *Boolean Functions and Equations*. North Holland, 1974.

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

[39] R. Sikorski. *Boolean Algebras*. Springer Verlag, 1964.

[40] P. Taylor. *Practical Foundations of Mathematics*. Cambridge University Press, 1999.

[41] L. Zhang and S. Malik. The quest for efficient Boolean satisfiability solvers. In Andrei Voronkov, editor, *Proceedings of CADE-18, 18th International Conference on Automated Deduction*, volume 2392 of *Lecture Notes in Computer Science*, pages 295–313. Springer, 2002.