Understanding One Another: Making Out AI Meanings with Boolean Equations*

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

November 20, 2005

Abstract

Autonomous AI agents raise the issue of semantic interoperability between independently architectured and differently embodied intelligences. This paper offers an apporach to the issue that is close in spirit to the way humans make out meanings. Using a mathematical model of cognition, it is shown how autonomously developed conceptualizations can bootstrap and unravel each other's meanings ad hoc. The domain general methodology is based on own Boolean capabilities, and any 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 conjectured conceptualization of the other, along with a testable conjectured translation that maps from that conceptualization to one's own.

1 Introduction

Autonomous AI agents raise the issue of semantic interoperability between independently architectured and differently embodied intelligences [1]. When intelligence emerges bottom-up from its embodiment and an own sensory-motor-neural aparatus, then the way each agent perceives the environment, and the ensuing conceptualizations and ontologies, could vary radically. The typical strategy of creating large, shared, conceptual schemas to be used as a common reference is problematic, because they are not grounded in the individual perceptions of the agents that access them. That typically requires pretailored inflexible translations to a fixed conceptualization of a database. This paper offers an apporach to the issue that is more adaptive, and closer in spirit to the way humans make out meanings. It is based on having communicating intelligences bootstrap and unravel meanings ad hoc. That is formalized by the construction and the solution of grounded Boolean equations.

This 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 rigrous mathematical basis. A category is defined where every intelligent state is a categorical object (with its own conceptualization), and morphisms are paths of commensuration inter- and intra- intelligences (this will be further elaborated in section 2). The issue that is addressed in this paper is about obtaining a translating morphism between different states (and their conceptualizations) when none is provided.

An essential virtue of mathematical modeling is, indeed, about abstractions and formal operations on abstract symbols. For that reason, the very idea of mathematical modeling often seems contradictory to the modeling of highly embodied, situated, agents, being and acting in a real tangible world, and deriving their intelligence from innately grounded symbols and representations. AI research

*To be presented at AIMATH06: Ninth International Symposium on Artificial Intelligence and Mathematics.

has been troubled, not to say plagued, by related controversies for a long time [20, 16, 11, 14, 2, 12, 3].

The underlying idea of ISAAC is that it is possible to find suitable abstractions that mathematically model precisely that: embodied agents, situated in, and reacting to, an outside environment, and deriving their intelligence bottom-up from innately grounded symbols and representations. The aim is a unified theoretical framework for AI, and AI artifact description, analysis, and development within that framework. Rather than developing a new mathematics, the goal is the deployment of known mathematical traditions towards a mathematical model, a unified ontology and language of discourse, and systematic AI implementations.

To make the presentation self contained, section 2 outlines the underlying formalism, summarizes foregoing constructions, results, and related interpretations of ISAAC's research agenda, that have already been published and presented ([8, 7, 5, 4] and other papers at the author's web site [6]).

Section 3 describes fresh import about a technique to make out meanings, that falls naturally out of the formalism.

2 Basics of ISAAC and its Foregoing Developments

2.1 The Category of Perceptions

The basic formalism was introduced in [8], along with the underlying pre-theoretical AI intuitions.

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* $\varrho : \mathcal{E} \times \mathcal{I} \to \{t, f, u\}$.[1]

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 a formal setting simple examples would include spacial coordinates, or *'the thing that sensor x currently attends to'*, 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. The three-valued predicate ϱ is the *Perception Predicate/p-predicate* that relates between w-elements and discriminations. (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 [18]. That would be determined by each one of the individual instantiations that are being abstracted.

Given $\mathcal{M} = \langle \mathcal{E}, \mathcal{I}, \varrho \rangle$ as a model of an embodied sensory-motor-neural apparatus, the development of the schema proceeds to formalize, bottom-up, what kinds of intelligence could be generated on top of that, to study the innate capablities that would be required for these generations, and ways to model that with rigor that can be scrutinized.

A commensuration between perceptions is formalized by *perception morphisms (p-morphisms)*:

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*:*

$$h : \mathcal{M}_1 \to \mathcal{M}_2 \tag{1}$$

consists of two set mappings:

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

There is a structure preservation condition, entitled No–Blur*:* $\forall \mathrm{w} \in \mathcal{E}$ *,* $\forall \alpha \in \mathcal{I}$ *, whenever* $\varrho_1(\mathrm{w}, \alpha) \neq u$ *then* $\varrho_2(h^{\mathcal{E}}(\mathrm{w}), h^{\mathcal{I}}(\alpha)) = \varrho_1(\mathrm{w}, \alpha)$.

[1] In [8] $\mathcal{E}$ was defined and used, but it was fixed, so $\mathcal{M}$ was defined just by the pair $\langle \mathcal{I}, \varrho \rangle$. In [5, 4] the definition was extended to $\langle \mathcal{E}, \mathcal{I}, \varrho \rangle$, with a variable $\mathcal{E}$.

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

Composition and the identity are defined by composition and identity of set mappings, and a theorem followed that perceptions with p-morphisms make a mathematical category, designated $\mathcal{PR}$. This provides a well developed mathematical infrastructure for a 'theory of perceptions'. P-morphisms and other categorical constructs have been applied to formally model quite a few cognitive processes. (Whenever there is no risk of misunderstanding, the superscripts of $h^{\mathcal{I}}$ and $h^{\mathcal{E}}$ may be omitted.)

From the evolutionary pressures point of view, reactions and behavior typically develop adaptively side by side with perception. The basic definitions 1 and 2 have been then extended to include reactions and behavior (papers at the author's web site [6]), and a fallout is the second *A* in ISAAC, which stands for 'affective'. That is a significant part of this formalism, but definitions 1 and 2 are sufficient for the current discussion.

2.2 Modeling Upscaled Perception with Boolean Constructs

2.2.1 Basics

Intelligence processes its sensory input using not only 'its eyes', but also 'its head', to make sense of its environment, yielding a conceptual system. One of the subsequent steps was hence the modeling of analytic cognitive processes and representations on top of this sense perception framework [7]. To model that, Boolean perceptions $\mathcal{M} = \langle \mathcal{E}, \mathcal{B}, \sigma \rangle$ in a subcategory $\mathcal{PR}^{\text{bl-I}}$ were defined. They 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. (e.g. the value of $\sigma(w, \alpha \vee \beta)$ is computed from the values of $\sigma(w, \alpha)$ and $\sigma(w, \beta)$ in an expected manner.) The categorical construction in [7] yields Lukasiewicz-style 3-valued truth tables [17], that may be applied algorithmically for that computation. P-morphisms in this subcategory are based on Boolean homomorphisms between the Boolean algebras of discriminations. A perception with a Boolean algebra of discriminations serves representational purposes and related procedural objectives:

- 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 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 [6] regards discriminations *as triggers of reactions*, and Boolean combinations of discriminations then serve purposes of upscaled integrative behavior. Moreover, in that context it is conjectured that the need to integrate simultaneous conflicting reactions in some sensible manner could be a significant evolutionary pressure behind the development of lattices of compound sensory-motor-neural discriminations.

- *Not the least significant feature of Boolean perceptions is that they allow Boolean equations for making out meanings, and that is the import which is introduced in section 3 of this paper.*

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 [5, 4] 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 latter is 'pro-synthetic' in that it takes cohesive w-elements as wholes that are basic building blocks, and maps between them. The former is 'pro-analytic' in that it 'slices' impressions of cohesive whole into separate discriminations as building blocks, and maps between them. Formally, a mapping is a mapping, so that the schematic construct looks the same. From the cognitive modeling point of view one gets mental processes that are 'connatural', or 'sibling', in a cetain sense.

This *duality* has further reaching 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 3.5.3 this will be done for the constructions of this paper as well.)
- In potential computational implementations, the same high-level architectural or computational modul, after having been generalized to work with different parameters, could be reused for 'sibling' processes: A modul that manipulates elements of $\mathcal{I}$ (of $\mathcal{E}$)) could also manipulate elements of $\mathcal{E}$ (of $\mathcal{I}$)), mutatis mutandis. Reusage of moduls is a recognized phenomenon in the natural context: Evolution theorists use the term *exaptations* [15] to refer to minor changes that make use of already existing capabilities to create new behaviours (where the significance of the capability naturally grows together with the number of behaviours that it supports).
- Taking this even a step further: Since the formalism proposes to model cognitive processes, this may suggest looking into the possiblity of different cognitive processes being based on the same undelying capability. Such 'sibling' cognitive processes are expected to emerge side by side. Even if they look superficially unrelated, they are indeed related in a deeper sense. (In section 3.5.3 this idea will be applied in the context of this paper.)

Following this idea, 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 modelling and analysis of compound analogies [5] and imaginative design [4]. In a structural sense, they are 'sibling' to analytic and interpretive representational processes.

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

- Upscaled analytic and interpretive representational cognitive processes.
- Upscaled compound analogies and imaginative design.

(The construction of these upscaled perceptions revealed a singularity that happens to model a known cognitive paradox.)

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* $\mathcal{PR}^{\text{bl-E}}$'.

2.2.2 Validity and Completeness

Notions of validity and completeness in the Boolean constructs model a certain sense of an optimal–under–the–circumstances conceptual system. These notions 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 $\mathrm{w} \in \mathcal{E}$:*

$$\begin{aligned} \varrho(\mathrm{w}, \alpha) = \mathrm{t} \quad &\Rightarrow \quad \varrho(\mathrm{w}, \beta) = \mathrm{t} \\ \varrho(\mathrm{w}, \beta) = f \quad &\Rightarrow \quad \varrho(\mathrm{w}, \alpha) = f \end{aligned} \tag{2}$$

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 $\alpha \in \mathcal{I}$:*

$$\begin{aligned} \varrho(w, \alpha) = t \quad &\Rightarrow \quad \varrho(z, \alpha) = t \\ \varrho(z, \alpha) = f \quad &\Rightarrow \quad \varrho(w, \alpha) = f \end{aligned} \tag{3}$$

The partial order $\trianglelefteq$ models perceptible lawlike patterns of (Boolean combinations of) constituents.

Defintions 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 [17], can be found in [7].

In a valid Boolean perception $\leq \subseteq \trianglelefteq$, and in a complete Boolean perception $\trianglelefteq \subseteq \leq$. As shown in [7], all Boolean perceptions are valid, but not necessarily complete. A typical Boolean perception would be somewhere in-between: some lawlike patterna 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 system.

2.2.3 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 cognitive transitions into perceptions that feature higher level capabilities, marrying the grounding provided by the embodied sensory-motor-neural apparatus with the advantages of 'The Laws of Thought' [9]. In [7], two canonical endofunctors of that type were defined and studied. A typical Boolean generation would lie somewhere in between the two. The most general one is described first.

2.2.4 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.
- 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 consituents and leading to a combinatorial disaster.

There is a functorial fixed point formalism that serves to characterize these drawbacks.

2.2.5 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 $\mathcal{G}^{\text{fr}}$ of section 2.2.4 above, and to enable the modeling of better perceptual acuity. Intuitively, one still wants to do a methodical open-minded general transition to a Boolean perception, but to somehow restrict it to those intelligences that show a 'natural tendency' to deal with a Boolean structure. Very loosely, this 'potential' turns out to demonstrate itself in the transitions (p-morphisms) that are allowed (e.g. the way one communicates) rather than in the perceptions themselves (e.g. one's embodiment).

The subcategory $\mathcal{PR}^{\text{Sk}}$, '*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 a certain type: *Sketch Structured p-morphisms* only[2]. The precise formal definition of the *sketched structure* is described in [7]. Loosely, this structure consists of traces of perceptible Boolean patterns, that may be present in non-Boolean perceptions as well. If, for example, $\mathcal{M}$ features discriminations {*infrared, visible, invisible, ultraviolet*}, then an acute perception should be able to observe that $\neg\textit{visible} \trianglelefteq \textit{invisible}$ and vice versa, also that $\textit{infrared} \vee \textit{ultraviolet} \trianglelefteq \textit{invisible}$, and so on. A sketch structured p-morphism h would preserve these patterns, so that $\neg h(\textit{visible}) \trianglelefteq h(\textit{invisible})$, and so on. (A similar idea of an intelligent analysis of data was independently developed in [10].)

[2] a subcategory with the same objects, but perhaps fewer morphisms, is sometimes called a *wide* or a *lluf* subcategory [22, p.211].

Another endofunctor $\mathcal{G}^{\text{fr-cmp}} : \mathcal{PR}^{\text{Sk}} \to \mathcal{PR}^{\text{bl}}$, which is a free functor that is defined on the sketch-structured subcategory of perceptions, yields a complete Boolean perception $\mathcal{G}^{\text{fr-cmp}}(\mathcal{M})$ (whereas $\mathcal{G}^{\text{fr}}(\mathcal{M})$ of section 2.2.4 above was, in the general case, incomplete). One then gets the further restricted Boolean subcategory $\mathcal{G}^{\text{fr-cmp}}(\mathcal{PR}^{\text{Sk}}) = \mathcal{PR}^{\text{bl-cmp}}$. The endofunctor $\mathcal{G}^{\text{fr-cmp}}$ is based on a construction that, loosely, 'moves things around' in the Boolean closure so that perceived patterns should be reflected by the target Boolean structure. Consequently, the perceptions in the subcategory $\mathcal{PR}^{\text{bl-cmp}}$, are *valid and complete* Boolean perceptions that may be generated over all basic perceptions $\mathcal{G}^{\text{fr-cmp}}(\mathcal{M})$.

The sketch structured generating p-morphism $\xi^{\text{Sk}} : \mathcal{M} \to \mathcal{G}^{\text{fr-cmp}}(\mathcal{M})$ is a natural transformation from the identity functor on $\mathcal{PR}^{\text{Sk}}$ to the functor $\mathcal{G}^{\text{fr-cmp}}$. This provides the modeling of a methodical cognitive transition from basic perceptions to Boolean perceptions that is still fairly open-minded and general[3], but this one is also *perceptually acute* because:

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

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

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.

3 Unravelling Meanings with Boolean Equations

3.1 The Problem

Section 2 showed how every perception has its own perceptual discrimination system and p-morphisms translate between different perceptions. Meaning is 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 'looking into each other's head'.

3.2 A Working Example

A traveller 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 traveller calls these forms: 'square', 'rectangle', 'rhombus'. The child utters 'boo', 'bla', and other unintelligible tokens. After a while the traveller 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 figure out 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 [13]. 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' and he undertook their formal investigation [9]. It is proposed that what the traveller was doing in that train can be modeled by solving a system of Boolean equations.

[3] 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.

3.3 Solving the Example

Let creative communication (or an inscription on a Rosetta Stone) 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:

$$\begin{array}{rcl} \mathit{rhombus} & \leq & \mathit{boo} \\ \mathit{rectangle} & \leq & \mathit{bla} \\ \mathit{square} & = & \mathit{boo} \wedge \mathit{bla} \end{array} \tag{4}$$

A detailed study of Boolean equations can be found in [19]. It analyzes how Boolean equations can be solved, algorithmically. Complexity issues are also dealt with. 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, including the following solution of (an equivalent of) the system (4):

$$\begin{array}{rcl} \mathit{boo} & = & \mathit{rhombus} \vee \mathit{square} \\ \mathit{bla} & = & \mathit{square} \vee \mathit{rectangle} \end{array} \tag{5}$$

The system of Boolean equations (4) happens inside a Boolean subalgebra of discriminations from the traveller's perception, where *square, rectangle, rhombus* are constant discriminations, and *boo, bla* are unknowns. The solution 5 provides discriminations from the traveller's perception as values for these unknowns. To wrap up, the traveller's perception features the generalizations:

$$\begin{array}{rcl} \mathit{rhombus} \vee \mathit{square} & = & \mathit{all\_sides\_equal} \\ \mathit{square} \vee \mathit{rectangle} & = & \mathit{right\_angles} \end{array} \tag{6}$$

For this, and other typical cases where unravelling meaning is called for, the perceived environment under consideration (e.g. the folded train tickets) is shared by the two sides.

3.4 The General Solution

First, one needs to formalize phrases like '*a rhombus is 'boo' (and perhaps more unintelligible things)*'. Definition 3 formalized lawlike patterns of discriminations $\alpha \trianglelefteq \beta$ in the context of one given perception. The essence of definition 3 is now abstracted to be used between two perceptions:

Definition 5 $\xrightarrow{\text{Luk}}$ *is a relation between truth values,* $\xrightarrow{\text{Luk}} \subset \{t, f, u\} \times \{t, f, u\}$*, given in table 1.*

Table 1: $\xrightarrow{\text{Luk}} \subset \{t, f, u\} \times \{t, f, u\}$

	t	f	u
t	+	-	-
f	+	+	+
u	+	-	+

As already said about definition 3, this is a Lukasiewicz-style 3-valued truth table [17], and $\xrightarrow{\text{Luk}}$ is non monotone, but without u values, it coincides with classical two valued material implication.

Definition 6 *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 Boolean perceptions. (Sharing the environment* $\mathcal{E}$*). A* 'Rosetta' *for* $\mathcal{M}_1$ *and* $\mathcal{M}_2$ *consists of* $n \geq 1$ 'observations' *of the form:*

$$\forall \mathrm{w} \in \mathcal{E} \quad \sigma_1(\mathrm{w}, f_\imath) \ \star_\imath \ \sigma_2(\mathrm{w}, g_\imath) \tag{7}$$

Where:

- $\star_\imath \in \{ =, \xrightarrow{\text{Luk}}, \xleftarrow{\text{Luk}} \}$.
- $\{f_\imath\}_{\imath=1,n}$ *are Boolean expressions in* $\mathcal{B}_1$.
- $\{g_\imath\}_{\imath=1,n}$ *are Boolean expressions in* $\mathcal{B}_2$.

Remark 1 *The treatment of this section works with all truth values and hence it will not be restricted. In practice, however, it does not seem sensible to make an observation just by virtue of things being undefined or false. That will add to the list of solutions too many 'theoretically possible' options that are not really justified. This will be discussed again in the fallout section 3.5.*

If $\mathcal{M}_1$ is the perception of the traveller, and $\mathcal{M}_2$ the perception of the local child, then the rosetta of the working example is:

$$\begin{array}{l} \forall \mathrm{w} \in \mathcal{E} \quad \sigma_1(\mathrm{w}, \mathit{rhombus}) \xrightarrow{\text{Luk}} \sigma_2(\mathrm{w}, \mathit{boo}) \\ \forall \mathrm{w} \in \mathcal{E} \quad \sigma_1(\mathrm{w}, \mathit{rectangle}) \xrightarrow{\text{Luk}} \sigma_2(\mathrm{w}, \mathit{bla}) \\ \forall \mathrm{w} \in \mathcal{E} \quad \sigma_1(\mathrm{w}, \mathit{square}) = \sigma_2(\mathrm{w}, \mathit{boo} \wedge \mathit{bla}) \end{array} \tag{8}$$

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

Definition 7 *Let* $\mathcal{M}_1$ *be a valid and complete Boolean perception. Assume a rosetta for* $\mathcal{M}_1$ *and*

$\mathcal{M}_2$. *The corresponding* 'system of equations' *is a system of n simultaneous Boolean equations in $\mathcal{B}_1$, which is obtained as follows: First, define φ by*

$$= \overset{\varphi}{\mapsto} = \quad , \quad \overset{\text{Luk}}{\longrightarrow} \overset{\varphi}{\mapsto} \leq \quad , \quad \overset{\text{Luk}}{\longleftarrow} \overset{\varphi}{\mapsto} \geq \tag{9}$$

Next, for all $1 \leq \imath \leq n$, an expression $\overline{g_\imath}$ in $\mathcal{B}_1$ is obtained from $g_\imath$ by replacing every constant $\beta \in \mathcal{B}_2$ of $g_\imath$ by an unknown variable $\overline{\beta}$ in $\mathcal{B}_1$. The i's Boolean equation is then written as: $f_\imath \ \varphi(\star_\imath) \ \overline{g_\imath}$.

Remark 2 *Definition 7 requires that $\mathcal{M}_1$ be a* complete *Boolean perception, so that observed patterns can be replaced by the Boolean partial order as in (9). This requirement will be somewhat relaxed in section 3.5.2.*

For example, the rosetta (8) gives (4) as its system of Boolean equations.

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

Theorem 1 *Let there be a rosetta for $\mathcal{M}_1, \mathcal{M}_2$ as in definition 6. 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 $\mathcal{B}_1$, 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 $h : \mathcal{B}'_2 \to \mathcal{B}'_1$, where $\mathcal{B}'_2 \subseteq \mathcal{B}_2$ is the subalgebra generated by the β's of the rosetta, and $\mathcal{B}'_1 \subseteq \mathcal{B}_1$ is the subalgebra generated by the h_β's.*

2. *There exists a perception $\widehat{\mathcal{M}'_2} = \langle \mathcal{E}, \mathcal{B}'_2, \widehat{\sigma_2} \rangle$ that is consistent with the observations of the rosetta, namely: $\forall w \in \mathcal{E} \quad \sigma_1(w, f_\imath) \quad \star_\imath \quad \widehat{\sigma_2}(w, g_\imath)$, and the homomorphism h defines a rigid p-morphism[4] $h : \widehat{\mathcal{M}'_2} \to \mathcal{M}'_1$.*

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

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

Proof.

1. To prove item 1 of the theorem: h could be extended to a Boolean homomorphism if and only if the following holds:[5] 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 $-1, 1$:[6]

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

implies that:

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

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 $\mathcal{M}_1$, the following can be added to the rosetta without loss of generality: $\forall w \in \mathcal{E} \ \sigma_1(w, \perp_{\mathcal{B}'_1}) = \sigma_2(w, \eta_1\beta_1 \wedge ... \wedge \eta_k\beta_k)$, and that adds the following to the system of equations: $\perp_{\mathcal{B}'_1} = \eta_1\overline{\beta_1} \wedge ... \wedge \eta_k\overline{\beta_k}$. 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: $h(\alpha) = \overline{\alpha}$. It follows that the corresponding 'source observations' of the form $\forall w \in \mathcal{E} \quad \widehat{\sigma_2}(w, \alpha) = \sigma_1(w, h(\alpha))$ are consistent with the rosetta. 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 corresponding 'source observations' of the form $\forall w \in \mathcal{E} \ \widehat{\sigma_2}(w, \alpha) = \sigma_1(w, h(\alpha))$ are *testable predictions* that could be applied to test the conjecture.

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

[5] See [21, p.36].

[6] Following [21] we use the notation $(-1)\alpha = \neg\alpha$, and $(+1)\alpha = \alpha$.

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 $\widehat{\sigma_2}(w, \alpha) \neq u$. 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.

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 $\forall w \in \mathcal{E}\ \ \sigma_2(w, \alpha) = \sigma_1(w, h(\alpha))$ for all (or just a set of generators) α in $\mathcal{B}'_2$. $\Box$

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 [8]). $\mathcal{U}_{\mathcal{E}}$ 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.

3.5 Fallout and Discussion

The methodology of ISAAC is to have results inferred from formal premises with context-free mathematical rigor, and that was meticulously (and tediously) done in section 3.4 above. However, 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. This is the essence of this section. 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.

3.5.1 Sorting the Solutions

What if there is more than one solution to the rosetta system? Following remark 1, 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.

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[7], for example a solution that involves the least complex Boolean expressions.

3.5.2 Relaxing Requirements

A look back at the construction shows that making out meanings as described above requires that the perception within which the system of equations is solved should be the 'more upscaled' perception (In the working example: the traveller). 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. However, some requirements may be slightly relaxed:

- Definition 7 requires that $\mathcal{M}_1$ be a *complete* Boolean perception, which does not sound realistic. (As already remarked in section 2.2.2, in a typical Boolean perception some lawlike patterna are observed and internalized, but not all.) However, by following the details of the usage of φ in that definition, this requirement may be relaxed: only the patterns that are observed in the rosetta need to be internalized in the Boolean structure of $\mathcal{M}_1$.

- The other perception, namely $\mathcal{M}_2$, is required to be a Boolean perception. However, by following the construction of the rosetta, and the

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

proof of theorem 1, it can be observed that it is actually the other intelligence that conceives of a Boolean generation over $\mathcal{M}_2$, in order to write down and to solve the equations within $\mathcal{M}_1$, but $\mathcal{M}_2$ does not have to be Boolean in itself.

3.5.3 Fallout by Symmetry

In section 2.2.1 a *duality* propety was described, based on the fact that the roles 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. It was shown here how to obtain a p-morphism which consists of an interpretive translation of the form $h^{\mathcal{I}} : \mathcal{I}_1 \to \mathcal{I}_2$ and $h^{\mathcal{E}}$ is the identity. (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 looking for analogies 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 could arise when an intelligence has to figure out an unknown environment. 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 [5].)

3.5.4 Some Issues for Further Investigation

A question that may be asked is about cases where both perceptions happen to be upscaled enough to do the job. In that case, can one get a significant improvement? Another possible question 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?

3.5.5 Conclusion

Using ISAAC, which is a mathematical model of autonomous cognition, it was shown how autonomously developed conceptualizations can bootstrap and unravel each other's meanings. The domain general methodology is based on own Boolean capabilities, and any 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 conjectured perception of the other, along with a testable conjectured translation that maps from that other perception to one's own.

References

[1] Meaning negotiation: Papers from the AAAI workshop. Technical Report WS-02-09, AAAI, 2002.

[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] 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.

[5] 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.

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

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

[8] 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.

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

[10] 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.

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

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

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

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

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

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

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

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

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

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

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

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