Self, Empathy, Manipulativity: Mathematical Connections between Higher Order Perception, Emotions, and Social Cognition

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

Abstract

Preserving one's own autonomous perspective within a society of other autonomoous agents, and integrating that with social behavior that is both sensitive and sensible, is an intricate task. This paper shows how ISAAC, a mathematical model of perceptual cognitive and affective capabilities, could integrate and deploy these capabilities to provide theoretical and computational support for formal agents that feature a sense of self together with social cognition and affect.

Introduction

Social cognition is often concerned with general cognitive issues, such as storage, access, and retrieval of information about the environment. The essence that sets social cognition apart are the existences that inhabit the relevant environment: animate agents having minds of their own. For social cognition, intentional objects of mind processes are hence also subjects of similarily structured mind processes, that, in turn, have their perceivers as intentional objects, and so on, ad infinitum. Each side needs to take that reciprocity into considertion: while one perceives, cognizes, emotes, and behaves, so do the intentional objects of its mind processes. In particular, they do that about each other.

Evolution naturally selected for humans a gift to relate to agents with mentality like themselves (Humphrey 1984). Example evidence comes from animism, notably in young children (Piaget 1926). They intuitively assign mental states to nonhumans. Autism provides example negative evidence from abnormal human behavior. The cognitive explanation of autism has become known as the 'theory of mind deficite', because autistics seem to lack the intuitive understanding that people have mental states (Baron-Cohen 1995).

Having to preserve one's own autonomous perspective, within a society of other autonomous agents, introduces inevitable controversies. Conflicting perspectives of different agents are sources of inter-agent friction, and also intra-agent difficulties, because harmonious coexistence with the environment is a self interest as well. The self is both a source of growth of the individual personality, and a source of social adaptation. Being unselfish would hardly be a virtue if one did not have a self, but an essence of the self is its selfishness. A familiar everyday example happens when a subject expresses anger at a fellow object for having done something wrong, and, along with being angry, the subject also feels compassion for the object, who is going through the unpleasant experience of being blamed. Bimodalities of similar nature are ubiquitous in affective social relationships. Here are a few examples: (i) Empathy, the ability to put oneself in another's place, is typically regarded as a gift. However, surrender of one's independent personality to identify with another to a point of infusion is objectionable. (ii) Pity is a strong feeling of sorrow for somebody, but it may also imply a tinge of satisfaction that the subject is in a better state. (iii) Expressions of feelings sometimes substitute the feelings themselves, as a matter of mere etiquette. $(iiii)$ Manipulativity is a salient social offense which is typically related to scheming methods that are being deployed to one's own advantage. However, justifiable intentions may also stand behind activities designed to influence others' behavior, and it has been argued that a significant evolutionary pressure behind intelligence is the need to manipulate fellow beings (Whiten & Byrne 1997).

All the above is indeed not new. The controversies have been noticed and expressed from antiquity: *'Thou shalt love thy neighbour as thyself'* (Leviticus 19,18) has been proposed as the encapsulation of all Biblical wisdom. However, the sober and non-conformistic Ecclesiastes rather suggests (in chapter 3): *To everything there is a season, and a time to every purpose under heaven ...* and the lines relevant to this context say: *a time to kill and a time to heal ... a time to embrace, and a time to refrain from embracing ... a time to love and a time to hate, a time of war and a time of peace*. A critical question facing social agents seems to be a sensible and sensitive demarcation of lines between the conflicting options, then acrobatic performance in order not to stumble, falling too often or too deep into the open traps on either side of the boundary, while exercising cooperative compassion (again...) to fellow acrobats around.

ISAAC (Integrated Schema for Affective Artificial Cognition) is a formal architecture that integrates cognition with affect (Arzi-Gonczarowski & Lehmann 1998a; 1998b; Arzi-Gonczarowski 1998; 1999a; 1999b; 2000a; 2000b; 2001). Deploying the tools of ISAAC, this paper studies possibilities for a formal modeling of affective social cognition as discussed above. The formal model is useful for:

- Grasping the underlying notions in precise theory-based terms.

- Providing computational support to model cognitive and affective social agents.
- Integrating social cognition and affect with general cognition and affect in a unified theory. This is significant in view of the fragmentation of AI research which has become an acknowledged problem of the domain.

A Pretheoretical Direction

At an intuitive level, it would be difficult to conceive of a zombie that is either empathetic or manipulative. This is different from the reactive social behavior that is exhibited by some organisms such as ants or bees. Although this author does not have a first person experience of being an ant or a bee, it is conjectured that their social behavior is not based on a 'theory of mind' of fellow ants or bees. It is unlikely that they feature emotional empathy or manipulativity. A zombie with a fixed repertoire of automatic 'one size fits all' reactions, extensive as it may be, still needs to activate reactions from its repertoire on the basis of clear discriminations between situations. It is conjectured that it will fail to discriminate social situations that involve internal happenings in the relevant agents. Such discriminations, and the resulting behavior, are not made on the basis of overt behavior (e.g. tears, smiles, and so on) but rather on some other kind of grasp of introvert events (in particular with other agents often tending to conceal their emotions). The neurological literature, for example, reports of episodes of automatism where the afflicted person was hurt even by inanimate objects (such as knives) without attending to its own injury (Damasio 1999). Could an agent that does not even attend to its own injuries have a grasp of others' emotions?

Hence, our pretheoretical intuitions, that guide the rest of the paper, yield that emotional social behavior is somehow related to consciousness. In order to conjecture about a connection between social sensitivity and consciousness, the following is observed:

- As argued in the introduction, social agents need to perceive, to emote to, and to cognize about, the perceptual cognitive and affective states of other agents.
- Without attempting to exactly define consciousness, it is generally agreed that conscious agents have self reflective capabilities: they perceive, emote to, and cognize about, their own perceptual cognitive and affective states.

Along with the introductory notes about social cognition, these observations provide motivation to study perceptions of perceptions. The case for perceptions of perceptions in general, in the context of ISAAC, is made in (Arzi-Gonczarowski 2001). It is conjectured here, for the specific purpose of social cognition, that a general capability to perceive, to emote to, and to cognize about, perceptual cognitive and affective states in general, could be the quintessence that also enables both social perception and consciousness. Following the goals set in the introduction, the next steps are about a formal model of perceiving agents, endowed with cognition and affect, that could make discriminations about perceptions in general.

Basic Formalism

To make the discussion somewhat self contained, this section provides an outline of ISAAC's most basic theoretical premises. (Advanced constructs and theorems are discussed in the cited works.) In the spirit of the biological evolutionary context, where lower level organisms are reactive, ISAAC boots the 'minds' of agents from their reaction driven perception. Various forms of higher level cognition and affect are structured on top of the basic perceptual apparatus. A *Perception* is defined as a 5-tuple[1]

$$\mathcal{P} = \langle \mathcal{E}, \mathcal{I}, \varrho\ \mathcal{R}, \mathcal{Z} \rangle$$

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

The set $\mathcal{E}$ represents a 'snapshot' of a perceived environment, a collection of environmental chunks, most typically objects or events, *world elements* (*w-elements*) that could perhaps be discerned by an agent. Even if the environment exists independent of its perception, its carving up into individuated w-elements typically depends on the agent. For example, where one perceives a single group of agents, another might perceive many individual agents.

The set $\mathcal{I}$ stands for the building blocks of discriminations, *connotations* that are internally afforded by the perceiving agent. They span the possible content of the representation and the distinctions that can eventually be made. For example, an agent that has the ability to discriminate face expressions would typically afford internal symbols that encode these expressions (e.g. *smiling, crying* and so on). Connotations could consist of alpha-numeric strings, but also of other data structures: icons, diagrams, pointers, calls to subprograms, etc.

The 3-valued *Perception Predicate* (*p-predicate*) ϱ relates w-elements and connotations, where t stands for definitely 'yes', f stands for definitely 'no', and u indicates that perception, for some reason, does not tell whether a certain w-element has a certain connotation. If w is another agent, then a perception could perhaps indicate that: $\varrho(w, smiling) = t$, $\varrho(w, crying) = f$, $\varrho(w, hungry) = u$.

$\mathcal{Z}$ is a set of conceivable behaviors for that perception. It is most likely to consist of things that the agent with that perception could really do: physical (e.g. move, fight, flight), mental (e.g. attend to something, reason, memorize, recall), overt (e.g. speech), introvert (e.g. change attitude). NULL is a legitimate element of $\mathcal{Z}$, standing for indifference.

$\mathcal{R}$ is the action tendency, emoting, function. Perception of a w-element w with a connotations α could conjure up the tendency for a reaction z. Examples: $\mathcal{R}(w, smiling, t)$ = SMILE models an agent that tends to smile whenever it perceives another smiling entity. $\mathcal{R}(w, smiling, f)$ = TELL_JOKE models an agent that tends to tell jokes whenever it perceives an agent that is not smiling. $\mathcal{R}(w, name, u)$ = ASK_FOR_NAME models an agent that wants to know the names of things. NULL is a default

[1] $\mathcal{R}$ and $\mathcal{Z}$ are new notations of concepts that have been applied from (Arzi-Gonczarowski 1998) and onwards.

reaction if nothing else is specified. $\mathcal{R}(-,-,-) \subset \mathcal{Z}$ is a set of those action tendencies that are not triggered by immediate external catalysts: actions towards general goals, dispositions, and so on. JUMP$\in \mathcal{R}(-,-,-)$, for example, models a jumpy agent.

Multiple, conflicting, emotions may be conjured simultanously, especially in social situations as discussed in the introduction. Such conflicts are one of the main pressures for the emergence of integrative higher level behavior, social behavior being, again, a salient example. One of the strengts of ISAAC is the modeling of perceptions that feature formal infrastructure for the integration and prioritizing of conflicting action tendencies. This is achieved through Boolean closures, constructs and theorems are discussed in the cited works.

Specific $\mathcal{E}, \mathcal{I}, \varrho, \mathcal{Z}$, and $\mathcal{R}$ provide a concrete perception. The mathematical objects $\mathcal{P}$ hence stand for basic embodied precognitions. They are high-level in the sense that they are presumed to layer on top, and be grounded by, a more basic apparatus, such as a neural network.

Environments, agents, and perceptions are dynamic and ever changing. A flow of perceptions, either inter- or intra- agent, is formalized by *perception morphisms (p-morphisms, arrows)*. Consider two perceptions:

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

A p-morphism $h : \mathcal{P}_1 \rightarrow \mathcal{P}_2$ consists of the set mappings:

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

With the structure preservation *No–Blur* condition: For all w in $\mathcal{E}$, and for all α in $\mathcal{I}$, whenever $\varrho_1(w, \alpha) \neq u$ then $\varrho_2(h(w), h(\alpha)) = \varrho_1(w, \alpha)$. The mapping $\mathcal{E}_1 \rightarrow \mathcal{E}_2$ is a formal tool to describe transitions in the perceived environment. For example, $\mathcal{E}_2$ may feature a new w-element, say an agent *w*, that is not perceived in $\mathcal{E}_1$, either because it just arrived, or because $\mathcal{P}_1$ does not attend to it, but $\mathcal{P}_2$ does: $\mathcal{E}_2 = \mathcal{E}_1 \cup \{w\}$. The mapping $\mathcal{I}_1 \rightarrow \mathcal{I}_2$ is a formal tool to describe transitions in the representation. For example, $\mathcal{I}_2$ may feature a new connotation, say *red eyes*, indicating an added distinction. Constituents (either w-elements or connotations) may be replaced or merged as well by the set maps. The values of the p-predicate may also be modified along arrows, as confined by the *No–Blur* condition, which binds the interpretive and the literal-analogical aspects, providing an analytic explanation to transitions between environments on one hand, and, on the other hand, grounding interpretational transitions in holistic experiences. (The arrowed representation of p-morphisms is not necessarily chronological. It models a relationship between two perceptions. Chronologically, the target perception could sometimes exist before the domain perception that features blurred perceptual values, deleted or split constituents, etc.) There are no general constraints on the transitions between the $\mathcal{Z}$-s and between the $\mathcal{R}$-s, allowing for maximum flexibility in the formalization of motivational transitions.

Technically, composition and the identity p-morphism are defined by composition and identity of set mappings, and perceptions with p-morphisms make a mathematical category, designated $\mathcal{P}rc$. Mathematical category theory provides a well developed mathematical infrastructure to capture the structural essence of perceptive cognition and affect, without being over deterministic. P-morphisms provide a versatile tool for modeling perceptual cognitive and affective transitions. The cited works applied specially trimmed p-morphism constructs, with equational reasoning involving relevant commutative diagrams, to model representation formation, imaginative design, analogy making, communication, learning, and more.

ISAAC is grounded by intuitions, but the treatment proceeds as if the semantic primitives were context independent. All definitions, constructions and results are tidily operated within the abstract mathematical framework. Constructions and results are examined with regard to our grounding intuitions, pre-theoretical conceptions, existing theories, and opinions about cognitive processes. Quite a few results that have not been anticipated at the outset provided supporting arguments that the proposed premises are probably adequate to model perceptual cognitive processes.

Perceptions of Perceptions

Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be perceptions as above, where each one of the two relevant agents having these perceptions is in the environment of the other. If $\mathcal{P}_1$ perceives the perceptual cognitive and affective state $\mathcal{P}_2$, makes discriminations about it, and emotes, then $\mathcal{P}_2$ is a w-element in the environment of $\mathcal{P}_1$, namely $\mathcal{P}_2 \in \mathcal{E}_1$.

Theoretical Worries

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

The main motivation to go ahead and formalize perceptions of perceptions anyhow, where $\mathcal{E}$ and $\mathcal{I}$ are allowed to be 'non classical' sets, as studied by (Aczel 1987), is that *these are precisely the theoretical difficulties, that are inherent in the construction, that model inherent perceptual cognitive difficulties, social dilemmas and deadlocks in particular.* Vicious circles (behavioral, emotional, and so on) do happen in social situations. A trivial example: two polite people aggressively insisting on yielding the right of way through a door. There are indeed more serious reciprocity

problems than that in social cognition and affect, as outlined in the introduction, and they need to be modeled. If a theoretical model of social cognition and affect had consisted of straight line computations that always converge, then that would have provided a major reason for serious worries concerning the validity of that model.

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

A perception that perceives itself, namely $\mathcal{P}_1 = \mathcal{P}_2$, is a special case of circular reference. It is hence not surprising that similar worries have been raised with respect to self reflection. An agent could recur into infinite regress, perceiving itself as it perceives itself, and so on, requiring more and more resources and eventually derailing the system. (Sloman 2000) classifies reflective emotions together with other perturbant states that involve partly losing control of thought processes. He also remarks that: *'Self-monitoring, self-evaluation, and self-control are all fallible. No System can have full access to all its internal states and processes, on pain of infinite regress'*. (Flanagan 1998) suggests that the role of consciousness may be regarded as '*...more one of interfering with cognitive processes that are designed to function well and generally do so...*'.

ISAAC offers to capture self reflection, with its unavoidable entanglements, and social cognition, with its unavoidable irksome deadlocks. They are both epiphenomenal on the *nonwellfounded* mechanism of *higher order perception*.

Higher Order Formalism

ISAAC is extended to model *higher order perceptions*: $\mathcal{P}rc$–categorical constructs are allowed to be w-elements in environments, designated as *higher order w-elements*. In addition to the abovementioned $\varrho_1([\mathcal{P}_2], \alpha)$ and $\varrho_1([\mathcal{P}_1], \alpha)$ (The square brackets are for mere reading convenience), other categorical structures are also considered:

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

Relevant discriminations are allowed as *higher order connotations*, modeling perceptions with a 'theory of mind'. Quite simply, higher order perceptions are allowed to apply the concepts of ISAAC just like the readers and the author of this paper: connotations, w-elements, p-predicate, emotive reactions, p-morphism transitions, and so on. Examples:

- The higher order discriminating connotation $[\varrho(w, \alpha) = x]$ models a *sense of perception*. The value of the higher order p-predicate $\varrho_1([\mathcal{P}_2], [\varrho_2(w, \alpha) = x]) = y$ means: *It is y (namely: t or f or u) that perception $\mathcal{P}_2$ perceives the w-element w as having/lacking/x...connotation α.*
- The higher order discriminating connotation $[\mathcal{R}(w, \varrho(w, \alpha)) = z]$ models a *sense of emotion*. The y value of the higher order p-predicate $\varrho_1([\mathcal{P}_2], [\mathcal{R}_2(w, \varrho_2(w, \alpha)) = \mathrm{z}]) = y$ means: *It is y (namely: t or f or u) that perception $\mathcal{P}_2$ has the emotive reaction z to perceiving $\varrho_2(w, \alpha)$.*

The sense of emotion would model the conscious feel of own emotions, when $\mathcal{P}_1 = \mathcal{P}_2$, or a grasp of another agent's emotions. A sense of self, and its boundaries, are naturally required to discriminate between the two. That could be systematized with higher order ownership connotations:

$$\varrho_1([\mathcal{P}_1], [myself]) = t \ , \ \varrho_1([\mathcal{P}_2], [myself]) = f$$

Higher order w-elements consisting of p-morphisms and other categorical constructs, model perception of transience:

$$\varrho_1([\mathcal{P}_2 \to \mathcal{P}_2'], [transition\_from\_yesterday]) = t$$

The above captures a sense of the transition of time, but senses of other transitions could also be modeled. Connotations may then classify the nature of the transition: change of location, change of interpretation, and so on. Compositions of multiple arrows may stand for longer biographies.

Perceptions of complex structures with arrows involve more then one, current, perception, and require memory and more symbolic representations. It would hence be legitimate to set them apart. For self reflection, when $\mathcal{P}_1 = \mathcal{P}_2$, this seems to be the distinction that (Damasio 1999) makes between *core consciousness* and *extended consciousness*. According to Damasio, core consciouness is the feeling of what happens when we see or hear or touch, that marks those images as ours, while extended consciousness provides the agent with a more elaborate sense of self, an identity, and places that identity at a point in individual historical time. The latter is modeled here by perception of arrows, and more examples are provided below.

Indirect Emotional Perception

Formal structures from the former section model direct perception of other agents: overt behavior can be directly observed, as well as detectable physiological reactions (Picard 1997). However, turbulent emotions can also be experienced behind a non revealing façade (such as an ordinary e-mail). It is suggestive that agents with a grasp of introvert emotions in others are often designated as 'perceptive' (although we are not dealing with direct observations) or as 'conscious' (although we are not dealing with self reflection). This kind of perception needs to be modeled yet. Pre-theoretically, the subject agent conceives of itself in the state of the object, and that conceived perspective yields participation in the object's emotions. To model that formally, a few tools from ISAAC need to be deployed. Most have been introduced in the cited works, and the novelty is in intertwining them as basic components that, together, yield social cognition and affect. The mathematical infrastructure models the intuitive pretheoretical feel in precise terms, providing computational support to model social agents:

- A formal model of analogy making between perceptions to project constituents from one perception onto another (Arzi-Gonczarowski 1999b).
- A formal model of how to join two perceptions, attending to the degree of interfusion (Arzi-Gonczarowski & Lehmann 1998b).

- An extension of ISAAC to support perceptions that relate to imagined situations and not just to the authentic, current situation. That was introduced in (Arzi-Gonczarowski 1999a), for purposes of imaginative design processes. The gift of empathy, a projection of a subjective state into an object, is indeed related to imagination.
- Higher order capabilities from the former section, first introduced in (Arzi-Gonczarowski 2001).
- When all the above is achieved, the joint higher order perception will eventually conjure emotive reactions to each one of the joined perspectives, the authentic and the conceived. They need to be integrated to yield an intelligent behavior. That is were the formal Boolean infrastructure for the integration and prioritizing of multiple action tendencies will be needed, as explained in (Arzi-Gonczarowski 1998).

Getting Involved

The least obliging form of joint perceptions is formalized by categorical coproducts (a variant where the environment is shared was studied in (Arzi-Gonczarowski & Lehmann 1998b, p.288-289)). It provides a rigorous model of a 'minimal change common expansion' - an expansion of perception to include other perceptions as well, with otherwise no modification. The coproduct yields a new perception that models a perception that joins the agent's own authentic perception with a conceived perception of the other agent, but with a clear distinction between the two, because constituents are suitably indexed, and there is no binding between foreign constituents. The following technical hints could be skipped by readers who are not interested in mathematical detail:
$\mathcal{P}_1 \oplus \mathcal{P}_2 = \langle \mathcal{E}_1 \oplus \mathcal{E}_2, \mathcal{I}_1 \oplus \mathcal{I}_2, \varrho_1 \oplus \varrho_2, \mathcal{Z}_1 \oplus \mathcal{Z}_2, \mathcal{R}_1 \oplus \mathcal{R}_2 \rangle$
The $\mathcal{E}_i$'s, $\mathcal{I}_i$'s, and $\mathcal{Z}_i$'s, yield set coproducts with the required injecting p-morphisms that index the constituents of the coproduct perception. The definition of the p-predicate $\varrho_1 \oplus \varrho_2$ for pairs $((w, \imath), (\alpha, \ ))$, where both coordinates originate in the same perception $\mathcal{P}_\imath$, that original perception is naturally preserved: $\varrho_1 \oplus \varrho_2((w, \imath), (\alpha, \ )) = \varrho_\imath(w, \alpha)$. This warrants that the injections (rigidly) stand the no-blur structure preservation condition. For pairs $((w, \imath), (\varrho, \ ))$ such that the coordinates originate in distinct perceptions, define $\varrho_1 \oplus \varrho_2((w, \imath), (\varrho, \ )) = u$. It is sensible not to bind constituents from distinct perceptions, and this is also required by the definition of coproducts. The general categorical construct hence resonates with our intuitions. The definition of the action tendency function $\mathcal{R}_1 \oplus \mathcal{R}_2$ for pairs $((w, \imath), (\alpha, \ ))$, where both coordinates originate in the same original perception $\mathcal{P}_\imath$, the original action tendencey is naturally preserved: $\mathcal{R}_1 \oplus \mathcal{R}_2((w, \imath), (\alpha, \ ), \varrho_1 \oplus \varrho_2((w, \imath), (\alpha, \ ))) = \mathcal{R}_\imath(w, \varrho_\imath(w, \alpha))$. For pairs $((w, \imath), (\varrho, \ ))$ such that the coordinates originate in distinct perceptions, the default null reaction would be least obliging: $\mathcal{R}_1 \oplus \mathcal{R}_2((w, \imath), (\varrho, \ ), u) =$NULL.

The joint perception could now apply higher order 'senses' from the former section, to capture a conscious sense of perception, and a conscious sense of emotion, to the conceived constituents as well.

Conscious Involvement

In a copoduct perception, let $\mathcal{P}_1 \oplus \mathcal{P}_2((w_1, 1), (\alpha_1, 1)) = t$ model the perspective of $\mathcal{P}_1$, and let $\mathcal{P}_1 \oplus \mathcal{P}_2((w_2, 2), (\alpha_2, 2)) = t$ model the perspective of $\mathcal{P}_2$, with the respective emotive reactions: $\mathcal{R}_1(w_1, \alpha_1, t) = \mathrm{z}_1$, $\mathcal{R}_2(w_2, \alpha_2, t) = \mathrm{z}_2$. Without loss of generality, assume that the injective perceptual cognitive and affective process took place in the agent whose original authentic perception was $\mathcal{P}_1$:

$$\mathcal{P}_1 \rightarrow \mathcal{P}_1 \oplus \mathcal{P}_2 \tag{1}$$

Hence z_1 is the original authentic emotive reaction of that agent, but now it is fused with z_2, possibly a very different emotive reaction. These are just tendencies, and an intelligent agent needs to figure out what it should really do.

- The two emotive reactions originate in two different perceptual 'slots'. A confused agent might try to effect both reactions. To avoid disordered emotional responses to multiple experiences, a mechanism of integration is required. Boolean constructs are deployed for that purpose, as explained in (Arzi-Gonczarowski 1998), providing theoretical and computational infrastructure for the integration of composite action tendencies.
- For a purposeful management and optimal integration of reactions to the advantage of the individual self, it would be useful if perception could sense where each one originates. However, the joint perception $\mathcal{P}_1 \oplus \mathcal{P}_2$ features symmetry between $\mathcal{P}_1$ and $\mathcal{P}_2$. That is where *extended consciousness*, a term coined by Damasio and mentioned already above, is required. A reflective perception of complex structures with arrows, where equation 1 is a perceptible w-element, could sort out the difference between the two emotive reactions z_1 and z_2. In simple words: 'z_1 is me, and z_2 is the other agent'. With extended consciousness, an agent is endowed with an elaborate sense of self that places the joint perception at a point in its own individual biography.

As argued in the introduction, manipulativity is the darker side of the same coin. A manipulative agent performs similarly structured perceptual cognitive and affective transitions to figure out the emotive reactions of the other, but the integration with its own perspective favors manipulation over participation. Manipulation can be effected, for example, by handling the environment (Arzi-Gonczarowski 1999a) to extract the required reaction from the other agent. When agents communicate, and are capable of conceived perceptions, a threat concerning an eventual manipulation of the environment could do the manipulative job.

Closer Involvement

A categorical *pushout* construction would model a natural process where fusion between the joint perceptions gradually increases. There could be constituents (w-elements, connotations, behaviors) in different perceptions that are essentially the same. These could be merged with a suitable p-morphism that models the relevant cognitive transition.

$$\mathcal{P}_1 \rightarrow \mathcal{P}_1 \oplus \mathcal{P}_2 \rightarrow PerceptionPushout \tag{2}$$

The intuitive idea behind p-pushouts is to enhance partnership in coproduct perceptions. Figuratively speaking, why have 'your connotation' and 'my connotation' when the discrimination is essentially the same, and why distinguish between w-elements that are equally discriminated by the two perceptions. Technically, it is a lax analog to set unions (as opposed to direct sums). A variant where the environment is shared was studied in (Arzi-Gonczarowski & Lehmann 1998b, p.295-296), formalizing shared meanings. The mechanism of intelligent and conscious involvement from the former subsection should generally apply here as well, if equation 2 is a perceptible w-element.

An additional advantage is more computational support for the association of overt behaviors with emotions. Assume that the connotations α_1 and α_2 from the former subsection are merged. If the behavior $\mathtt{Z}_2$ is observed, then it is automatically associated with the merged connotation, and not only with α_2, as would have been the case in the more general coproduct. The shared meaning enables a more involved interpretation of behavior.

Provisions of Analogy

A prudent application of emotional social involvement should be conducted with a reservation, because a perception that perceives another perception always does that from its own perspective. A transition from $\mathcal{P}_1$ into a joint perception has been formalized in equation 1, but who provides the details of $\mathcal{P}_2$? Communication with, and observation of, the other agent may help, but are typically not enough. Perceptions can be extended only to conceived situations that they can at all represent, this was formalized within ISAAC in (Arzi-Gonczarowski 1999a). In humans, the missing constituents of another perception are filled in by analogies, metaphors, and similes from the perceiver's own subjective perception, from other agents, and from stereotypes. The underlying process is a process of analogy making and generation of analogical schemas. This was formalized within ISAAC in (Arzi-Gonczarowski 1999b). However, it is not always the case that if agents are alike in some ways, they will be alike in others, and it is not always possible to show that the resemblances noted bear relevantly on the points to be established. Indeed, empathetic involvement sometimes misses its point, with the subject projecting on the object conceived imaginary emotions that are not at all experienced by the object. For similar reasons, manipulativity does not always succeed.

Summary

In ISAAC, complex structures stand for higher level intelligent processes that are being modeled. Emotional social sensitivity may take a lifetime to tune and integrate with a refined sense of self. It is hence not surprising that a non trivial combination of formal constructs has been deployed to model intelligent empathetic (or manipulative) formal agents. Not the least advantage of ISAAC is that all these fragments of perceptive cognition and affect can be combined in a rigorous and general manner, being all based on the same theoretical premises, and yielding a unified theory.

References

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

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

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

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

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

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

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

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

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

Baron-Cohen, S. 1995. *Mindblindness*. MIT Bradford.

Damasio, A. 1999. *The Feeling of What Happens*. Harcourt Brace & Company.

Flanagan, O. 1998. Consciousness. In *A Companion to Cognitive Science*. Blackwell. chapter 9, 176–185.

Humphrey, N. 1984. *Consciousness Regained*. Oxford University Press.

Piaget, J. 1926. *La Representation du Monde chez l'Enfant*. Paris: Presses Universitaires de France.

Picard, R. 1997. *Affective Computing*. The MIT Press.

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

Whiten, A., and Byrne, R., eds. 1997. *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge University Press.