From Embodiments Back to their Models: An Affective Abstraction

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

Abstract

At the intersection of cognitive science and robotics are, among others, implementational issues that entail insights into the nature of embodied intelligence. This paper discusses one such example issue. It is argued that by explicitly entering physical robustness and perseverance of action into a formal cognitive theory, a wide range of behavioral, emotional, and affective phenomena can be inferred. In that case they are intertwined naturally with other intelligent phenomena, not as a shallow artificial patch, but rather as an essential part of a surviving intelligence.

Introduction

When putting cognitive science and embodied robotic systems side by side, the former could take, in a certain sense, the role of setting the general theory, while the latter would be about a variety of embodied instantiations of that general theory. From that viewpoint, cycles of feedback should ideally follow, as lessons learned from producing and observing instantiations, and grappling with their details, cause adjustments and changes in the theory, with repeated iterations of modifications to both the theory and to its implementations. Such back and forth adjustments are at the core of any creative process. In a sense, the original instantiations to work on were natural intelligences that (the predecessors of) cognitive science set out to explore, to introspect, and to model.

A distal ambitious goal of the co-evolution of a theory and its instantiations is to achieve the best possible two-way fit: On one hand, the full range of intelligent capabilities, that cognitive science has been able to model, should hopefully be implemented in robots one distant day. On the other hand, emerging, recurrent, details of implementations could be significant. If recurring implementation phenomena appear to be necessary and shared by all instantiations, then they are constituents in the intersection of cognitive science and robotics: Their essence should be properly extracted and embedded into cognitive models, supplementing the theory as the shared necessary essence of all its instantiations.

Both directions of this feedback loop between cognitive science and robotics are challenging, but the latter seems to be more evasive and it received less attention, at least until ideas about *embodied cognition* (Anderson 2003a; Chrisley 2003; Anderson 2003b) came up, with (Brooks 1999) being one of the leaders. Researchers are of course interested in implementing high-level intelligence in robots, and they keep working towards that and testing the results. When it comes to the opposite projection, significant details are sometimes being 'left for an implementation', overlooked, by both sides, as pure technical matters, and insights risk being lost. If that happens, then cognitive models would remain missing, flawed, incomplete.

This paper discusses an example where salient aspects of intelligence could be better integrated into formal cognitive theories if certain obvious implementation details made their way into the cognitive model.

Background

This work applies standard, well known, mathematical concepts, though this is *not* a technical mathematical paper. From antiquity, mathematical theories have emerged as formal modeling tools to rigorously describe given domains: Counting, mensuration, and calculation modeled the domain of quantities and operations on them; Geometry described physical space. These are but two examples.

In the context of mathematical theories serving as rigorous and general descriptive tools, two natural concepts that emerged are *soundness* and *completeness*. Very loosely rephrased: A theory is *sound* if anything that it either explicitly assumes, or adequately deduces from its explicit assumptions, holds in the domain that is being described; A theory is *complete* if anything that holds in the domain that it describes is either explicitly assumed by the theory, or can be adequately deduced from its explicit assumptions. A related famous example from mathematical ethos is about Euclide postponing the application of the parallel postulate[1] as long as possible in his *Elements*, and then finally coming to the conclusion that he had to explicitly assume that as an axiom: That property of our physical space is intuitively obvious, yet it could not be deduced from his other postulates, and was hence required as an explicit postulate to continue his theoretical undertaking.

[1] Given a line AB and a point P not on that line, exactly one parallel to AB can be drawn through P.

Modern, 'pure', mathematical theories describe domains that are themselves results of abstractions that require interpretation (and that earned the field the dubious reputation of detachment). Mathematical theories, either 'pure' and abstract, or directly grounded in concrete scientific research, are expected to be sound with respect to (some idealization of the current knowledge of) the domains that they describe, but they are more than often incomplete.

To warrant that formal properties can at all be studied and verified, mathematical practice makes a point of ignoring the intuitive grounding interpretation of formal symbols, and treating them as if they were meaningless. That way, all self-evident premises, as well as other assumptions, must be stated explicitly as axioms (postulates), and no hidden stipulations enter 'by the back door' from the pre-theoretical intuitions. Only when a theoretical result is formally deduced, it is interpreted into the grounding domain. (Mathematicians' irritating habit to make it hard on everybody, including themselves, by using meaningless symbols, is, in a sense, an inversion of the grounding problem.)

These background ideas are relevant to the interrelation between cognitive science and embodied systems such as mobile robots and autonomous vehicles, though the related areas of research are less formal, and not as scientifically mature: On one hand, while embodied robotic systems are still far from implementing all the features of a human level cognitive theory, one would naturally like that, at least, instantiations should not feature obvious contradictions of the underlying cognitive theory. That 'cognitive plausibility' could be regarded as a weak, lax, version of soundness. On the other hand, whenever one notices properties that seem to be necessary and shared by all instantiations, then one should try to somehow embed them into the cognitive model, as a small step in the general distal direction of completeness. The idea of proceeding in that direction, provides background for the wishful idea from the introduction, that cognitive models should seriously try to gain insights from robotic implementations.

The above raises another issue, that is also related to mathematical practice. Robotic implementations are, by their grounded nature, detailed, specific, and deterministic. They vary from one platform to another, and if there are underlying similarities between them, then they are not necessarily obvious. Observation of shared, recurrent, details of implementations would often depend on: ($\imath$) A series of suitable generalizations and abstractions, that might, if at all, eventually reveal an underlying shared essence, and ($\imath\imath$) A readiness to choose and to part, along the abstraction process, from discriminations and distinctions that may have been extremely significant and crucial at the implementation level. Also, needless to say, the abstraction process should not continue forever, lest one would get totally lost in clouds of over generalization that blur all meanings.

Indeed, resolving collections of phenomena into theoretical units of abstraction, modeling them with shared concepts, has always been a hallmark of scientific visions, and it is not an easy task. This brings us back to mathematical thinking, that is essentially about abstraction (Devlin 2001; 2003).

Physical Robustness and Perseverance, For Example

A Pre-Theoretical Fact

An implementation of intelligence in a robotic system, that interacts with its environment, is unlikely to work without physical robustness of the construction and its performance. An open ended diversity of environmental phenomena might endanger any embodiment, and an open ended diversity of obstacles might prevent actions of that embodiment from being consummated. A robot that lingers, perhaps even falls apart, if the wind is in an unfavourable direction, or the terrain in an unfavourable slope, is unlikely to survive. In nature, this issue is sometimes solved 'by big numbers': given zillions of feeble organisms, then, statistically, a certain percentage of them occasionally survives, and that might be enough to warrant survival of a species. At the level of single systems other solutions are required. If high-level intelligences were always able to fully anticipate, correctly, efficiently, thoroughly, and on time, all their upcoming oppositions and obstacles, and invariably come up with solutions that let all their activities be consummated and keep them intact, that might have been an answer to the problem. However, since all that is unlikely, embodiments simply need to be reasonably robust. At the level of single systems, survival depends on sturdy embodiments and perseverant actions: Their excitation should try to persist through thick and thin, and overcome eventual obstacles. That embodiments should not be made of, say, paper foldings, is self evident, perhaps too evident.

In the spirit of the introductory and background considerations, this is the kind of recurrent detail of embodiments that appears to be necessary and shared by all implementations, and hence its essence should be properly extracted and embedded into the theory, thus advancing the theory towards a more comprehensive description of its domain. Moreover, it will be argued that extending the theory with a suitable abstraction of that detail leads to far reaching theoretical inferences that may further improve a cognitive model.

Abstraction of the Fact as a Theoretical Postulate

The purpose of physical robustness and perseverance of action is to make it hard to interfere with an agent's state and with whatever it is doing. In other words, considerable resources (e.g. force, energy, power) should be required for such interferences. For mechanical activities these are the basic ideas of Newtonian mechanics, and similar principles apply for electrical, chemical, and other activities as well. Robust and perseverant agents feature inertia, and their state cannot be simply and instantly changed, or turned on and off (as opposed to bits in computer simulations). These basic properties of physical robustness and perseverance of action should be non specific and incontingent. Their essence is to resist an open ended diversity of oppositions and obstacles, hence they should always be present. The embodiment of an agent is indeed a given datum that cannot be easily changed. Actions/reactions of an agent are typically being triggered for reasons that have to do with that agent's concerns, and they determine the perseverance of the relevant activity (whether the concerns are conscious or not, and the related activities intended or not).

It is well known (though rarely explicitly stated in cognitive models) that activities of embodied systems require resources (e.g. power supply, maintenance) and these resources are applied along time to build up, to accelerate, and to maintain these activities. To add physical robustness and perseverance of action as a theoretical postulate, it is noted that it would obviously also take resources, applied along time, to derrange the state of an agent, or to deccelerate and to dissolve an activity once it has been triggered.

Formally, let us examine the option of associating with every possible action, say z, of an embodied system, a *Vitality Value* $V(z)$, to be evaluated by (a combination of) an open ended diversity of resources: power, tear-and-wear, and so on. Rather than modeling the resource consumption for the execution of the relevant action, this models the resources that should be invested in its obstruction (though the two may of course be related)[2]. Consider 'a postulate of physical robustness and perseverance', along the following lines:

Let z be an action. Once z has been triggerred, an amount V(z), to be evaluated in resources, is invariably required for its arrest.

We humans actually use physical and technological metaphors to describe our activities and their dissolution as well: heat, stir, fire, steam, shake, then let off steam, cool down, and so on. For artificial embodiments, tension, pressure, stress, or strain, may be literal, also being worn out, battered, and shattered. In this sense, moving from the natural to the artificial context is a reversion from everyday metaphors back to literal meanings.

Theoretical Inferences from the Postulate

The introduction of a postulate of physical robustness and perseverance into a cognitive model, using vitality values $V(z)$, makes it possible to infer a wide range of phenomena in a general and rigorous (i.e. *formal*) manner.

The suggested premise is intended primarily, as argued above, to model protection from interference or obstruction by external causes, either unintended (e.g. a strong wind, a steep slope), or intended (e.g. opponent agents in competition). That is also the primary evolutionary pressure behind natural robustness. A struggle between an agent and an external obstacle, or opponent, is typically observable from the outside. However, since a postulate of physical robustness and perseverance, as argued above as well, needs to be always present, non specific, and incontingent, then it has to be taken into account *also when inside mechanisms, within that very agent, try to arrest, deccelerate, or dissolve, an action/reaction that has been triggered within the agent itself.* In these cases the arena of action is introvert and not necessarily observable from the outside.

[2]For activities that are not vital and could be easily stopped, the vitality value $V(z)$ would indeed be set to a very low value, even to *null*, capturing neutrality, indifference, or apathy.

Internal Struggle, Literally

Environments often feature multiple stimuli, triggering multiple reactions simultaneously. If, technically, these actions cannot be performed together, a system risks confusion with disordered responses. The paradigm conflict is 'fight or flight?' that might cause freezing, which is hardly a recommended reaction in danger situations[3]. There is hence a natural evolutionary pressure to develop mechanisms that deal successfully with multiple, conflicting, reactions, salvaging disordered responses and replacing them with something more adequate for the concerns of the system.

One might have suggested conditioning the very activation of reactions (e.g. `perform(x) unless`...), but that is likely to turn out too cumbersome in emergencies. It has already been argued above that thorough anticipation of such a (possibly unpredictable, even endless) list of conditions and reservations is probably impossible. Rather than calling off basic reactions in advance, a more practical option would be to let reactions be triggerred unconditionally, and as conflicts occur, call for higher order control and arbitration mechanisms, so that the system's behaviour that is finally and actually generated, should be adequate.

Prioritization and selection are perhaps the simplest option for control and arbitration: one reaction takes precedence, while other reactions are arrested (e.g. react to the closest danger and stop all other actions). A more complex and creative possibility would be to substitute or integrate elements from a few reactions into a single coherent behaviour that perhaps compromises a little, but takes care of (almost) everything. (Decision making about actions to be taken constitutes an emotional capability à la (Damasio 1994).) *A shared property of all these mechanisms is that they involve a (partial) arrest of reactions that have already been triggerred, and that is where the new postulate comes in:* Legitimately perseverant impulses have been triggerred, and they might persist while being obstructed and forced to fade out by just as legitimate selection/integration mechanisms. Resources are being invested in maintaining and containing both sides of that arduous process. Internal mechanisms may be active just to cancel each other. Locally, this may not seem rational, but this could be the result of a globally sensible design as just explained.

Instead of actions and reactions that are being invariably consummated once they have been triggerred, one gets *action tendencies* that are being conjured up, but may eventually be arrested, and not necessarily executed, for a variety of reasons. Given an agent that features internal mechanisms to control its own responses, the postulate of physical robustness and perseverance has just provided us with a formal theoretical basis for the modeling of emotions à la (Frijda 1986), who claimed that the core of an emotion is a readiness to act in a certain way. The vitality value $V(z)$ may allow capturing that computationally.

Consider, for instance, an agent whose task is to guard a place, where only authorized agents are allowed to enter.

[3]Birds in 'fight or flight?' situations, for example, may preen their feathers as a *displacement activity* (Isaacs, Dainntith, & Martin 1996, p.217), jeopardizing their survival.

The guard is expected to be ready, at any time, to physically block the entrance of unauthorized agents who might try to get in, and to keep the entrance clear, possibly by pushing intruders back. When someone approaches, then in a human context we would have said that one needs to be 'on guard': attentive and alert, ready to react, perhaps even take a preparatory step forward. In engineered contexts one would have said that the engine needs to be hot and running, perhaps even start to move to gain impetus. If that turns out to be an authorized visitor after all, there is action tendency only, and the action itself is called off. The guard did not do much but resources have been consumed and introvert frictions did occur.

Needless to say, not all conflicts betweem actions or reactions have solutions, but even if there exists a solution to a conflict as described, then the required resources may not be available, especially if that involves arrest of reactions with high vitality. It may hence be hard to effectuate the solution. That provides a theoretical computational basis for the modeling of tertiary emotions à la (Sloman 2004). Even evolution did not manage to tune amounts of perseverant vitality that are *just right* for an effective management of conflicts in all possible scenarios. Legitimate action tendencies occasionally resist legitimate control, derailing mind function. (Imagine, for instance, the high vitality of the guard's reactions if its own existence had also been involved, and the visitor had been provocative, or slow in showing its authorization.) In humans, behavioural balance is indeed a non-trivial mind feature that is inherently wobbly, and generates a wide range of both overt and introvert phenomena. In robots, behavioral conflicts would not generate the human visceral experience of emotions, but rather (a variety of) resource consumptions, also tear-and-wear, to arbitrate or integrate between conflicting reactions with high vitality $V(z)$ values. If that does not succeed, then the robot is likely to exhibit a behavior that, in a sense, would be the analog expression of strong human emotions that cannot be contained.

Beyond Immediate Survival

Cognitive models of higher level intelligence feature more than a plain account of environmental stimuli and immediate responses. Cognitive science and AI have been extensively studying ways in which high level intelligences feature mental activities involving memory, internal representations, operations on these representations such as planning and problem solving, and so on. The evolutionary pressure behind these activities remains the same: they are there because they happened to improve chances of survival. An agent may react to a situation, for example, by retrieving memorized analogical situations, in order to apply solutions that proved successful in those similar situations. In other cases, given time and other computational resources, an agent may use internal representations of situations to compute, 'offline', better solutions that are too complex to be devised ad-hoc. (For instance, the guard may be recalling previous times when it had to deal with intruders, trying to learn from that, and planning an improved reaction in the future.)

Assuming that these upscaled capabilities evolved gradually in small steps, it is highly likely that they 'ride' on the original, basic, stimulus-response circuitry[4]. When an upscaled mental activity (e.g. memory, planning, or communication activities) z is triggerred using the original, basic, stimulus-response circuitry, it would also come with a vitality value $V(z)$, and that activity would possibly have to compete with other conflicting activities over resources and over precedence. Put together with a postulate of physical robustness and perseverance, that provides us with a theoretical basis for the modeling of emotions that are triggered by mental action tendencies. In the human context, thought processes, for example, may have their own vitality and perseverance. A person who is interrupted in the course of a thought process might have an emotional attitude towards the interruption. It has just been shown how that may be inferred formally.

When applying the stimulus-response circuitry to a representation, which is a conceived, rather than authentic, situation, then that would trigger action tendencies that, of course, should not be consummated, because the situation is just hypothetical. Here, again, a postulate of physical robustness and perseverance participates in providing a theoretical basis for the modeling of additional emotional phenomena: 'as if', deliberative, emotions à la (Sloman 2004).

Internal Resource Management

Given a postulate of physical robustness and perseverance, both overt and introvert activities have been inferred, that consume resources. Resources that wane, risk draining out unless timely recharged. Recharging batteries may be straight forward, but perseverance could also cause injuries, such as effects of friction, or springs that might deflect beyond their elastic limit. In both artificial and natural contexts there may be effects of both external and internal exhaustion, that may or may not be repairable. (Imagine, for instance, the exhaustion of the guard after numerous alerts.) Example mechanisms that recharge vitality resources could be ($\imath$) Automatic regeneration paced along time, abstracting routine maintenance or cooling down in machines, and 'rest', 'recuperation', or 'recovery' in humans. ($\imath\imath$) Some reactions, also effects of successfully consummated behaviours, could consist of an increase in vitality, roughly modeling the effect of natural 'contentment' or 'satisfaction'. Whether there could be a phenomenal analog for that in machines is an interesting question. The various resources involved may or may not be converted, reduced, or sublimed, one to the other. (In the natural context, a meal, also a good fight, could be rewarding in more than one sense, thus hunger would be depriving in more than one sense.)

Self Perception Phenomena

A high level system may feature some perception of its own mechanisms. The evolutionalry pressures behind that are probably the ones that endowed humans with self reflection. awareness, and other aspects of conscious beings. When a robust and perseverant agent features a perception

[4]Apart from cognitive plausibility, there are additional advantages in a small set of building blocks for engineered artificial agents (Lipson, Antonsson, & Koza 2003).

of its own perseverance and vitality values, its vitality resource consumption, and its overall internal resource management, then it may sense, for example, that certain relevant resources have been, or are going to be, drained below a safety margin. The system may hence be going to experience difficulties in continuing to produce a sensible integrative behaviour (i.e. it is going 'to get out of its mind'). That agent may hence take high level control actions to deal with that, including giving up, diverting to activities that use more available resources, taking a maintenance rest, and so on. (That, too, may be either conscious and intended, or not.) A postulate of physical robustness and perseverance thus participates also in providing a theoretical basis for the modeling of higher level behavioral control.

Social Cognition

A high level system that features some perception of its own mechanisms, may eventually be able to infer the mechanisms of fellow agents, especially if they are similar to itself[5]. In addition to reasoning about patterns of stimulus-response in fellow agents, it may also be able to reason and infer their perseverance and vitality values, the ensuing internal struggles, their vitality resource consumption, and their overall internal resource management. That opens options for phenomenal empathy, as well as for scheming manipulativity. A postulate of physical robustness and perseverance thus participates also in providing a theoretical basis for the modeling of higher level social cognition and understanding of other agents.

An Example Application

A postulate of physical robustness and perseverance, as suggested above, has been embedded into ISAAC ('Integrated Schema for Affective Artificial Cognition'). ISAAC is a mathematical model of intelligence, which gives rise to a formal theory that could be implemented computationally (Arzi-Gonczarowski & Lehmann 1998b; 1998a; Arzi-Gonczarowski 1999b; 1998; 1999a; 2000a; 2000b; 2001a; 2001b; 2002; 2003; 2004).

Similar to the natural evolutionary context, the schema starts from a simple model of corresponding sensations and reactions. It then structures 'upgrades' (e.g. handle conflicting reactions, internal representation, and so on) on top of that, using generative reasoning to systematically obtain and study the properties of these upgraded structures. Among other things, this approach models a continuous bridge from low level to high level intelligence.

A *perception* snapshot is structured as a set of *world elements* that constitutes an environment (real or imagined), a set of *connotations* that constitutes a collection of discriminations, and a set of *behaviors*. Behaviors are conjured up on the basis of a *perception predicate* that relates between world elements and their connotations in a three valued manner (*true, false, undefined*). With real environments, that basic schema approximates models of simple forms of intelligence. For higher level forms of intelligence, mind activities (cognitive, behavioral, affective) are modeled as streams of perceptions. Along these streams, all the components mentioned above could adapt dynamically: be modified, extended, merged, and so on.

In the course of a few years of ongoing research, a variety of mind processes have been modeled on the basis of these uniform, yet flexible, premises, capturing mental activities from streams of interpretations, through behavior development and integration, representation formation, imaginative design and anticipation, analogy making, to social and self perception. In addition to modeling single mind aspects, the collection features an additional value of an integrated whole: because they share uniform modeling premises, the various processes can be neatly composed and alternated between, modeling multifaceted intelligences.

Results, that have not been anticipated at the outset of ISAAC, provide supporting arguments that the proposal is apparently on a promising track, towards a unifying theory and, hopefully one day, ($\imath$) Cognitive science becoming a science with a sound formalism, and ($\imath\imath$) Artificial intelligence that integrates embodiment with high level cognitive and affective processes, not losing the big picture by over fragmentation.

The postulate of physical robustness and perseverance has extended the formal framework of ISAAC in the general direction of a sound theoretical description of behavioral, emotional, and affective phenomena as discussed above.

Methodological Issues

Philosophical issues, related to reductionism and supervenience, are outside the author's expertise. The essence of mathematical modeling has always been to start from basic concepts that are intuitively convincing and obvious. Then, following a consecution of simple steps, each one intuitively convincing by itself, one obtains arbitrarily complex and high-level constructs. A typical paradigm is the system of natural numbers: The five postulates of Peano capture the pre-theoretical essence of the natural numbers as counters. These simple and intuitive premises provide basis for complex high-level results that are far from obvious (e.g. Fermat's last theorem). Orderly extensions to the natural numbers provide the integers, then the rational numbers, then the real numbers[6].

A vitality value, $V(z)$, that is associated with each and every action or reaction of an agent, captures an integrated combination of two basic intuitions: ($\imath$) Actions triggered by the system are invariably legitimate, important, and vital, even in cases when they happen to be obstructed or called off for rational and other legitimate reasons, and ($\imath\imath$) This embodied system is a robust, surviving system.

Like a reduced instruction set for a computer, a systems approach conflates the types of building blocks that are required for an architecture, but not necessarily the spectrum of phenomena that are thus modeled. Because the

[5]An underlying idea of intersecting cognitive science with robotics is that, at some level of abstraction, AI agents and humans are similar.

[6]Hilbert coined for that the term *The Genetic Method*, which is suggestive in the context of biological complexity.

model is bootstrapped from basics, and proceeds heel-and-toe, one gains insights beyond shallow patterns of behavior, but rather a deeper understanding of what drives that behavior, which is a wide range of internal phenomena.

Summary

When putting cognitive science and embodied robotic systems side by side, a scientific bimodality emerges. On one hand one is trying to approximate intelligence by creating particular models of intelligence, such as mobile robots and autonomous vehicles; On the other hand, one is trying to formulate theoretical foundations for a general account of intelligence. The theory should not be dodging the embodied grounding issue, but rather providing tools of rigour that capture the essence of all its instantiations.

As an example, it was shown how a postulate of physical robustness and perseverance, when embedded into a cognitive model, and combined together with other capacities, has a potential of enhancing the model towards a sound formal description of a wide range of behavioral, emotional, and affective phenomena in a general and rigorous manner.

By entering perseverant vitality into a cognitive model explicitly, by the front door, emotions and affect follow naturally, intertwined in a theory of intelligence, not as a patch or as an artificial addition to mimic human behavior in a shallow manner, but as an essential part that emerges from an embodied, surviving, intelligence.

References

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

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

Arzi-Gonczarowski, Z., 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. S., and Maher, M. L., 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. 2001a. Perceptions that perceive themselves – a mathematical schema. *IJCAS: International Journal of Computing Anticipatory Systems* 8:33–51.

Arzi-Gonczarowski, Z. 2001b. Self, empathy, manipulativity: Mathematical connections between higher order perception, emotions, and social cognition. In Cañamero, D., ed., *Papers from the 2001* AAAI *Fall Symposium: Emotional and Intelligent II – The Tangled Knot of Social Cognition*, 9–14.

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

Arzi-Gonczarowski, Z. 2003. Computational synthesis: Following the treaded mathematical track. In Lipson, H.; Antonsson, E. K.; and Koza, J. R., eds., *Papers from the 2003* AAAI *Spring Symposium on Computational Synthesis: From Basic Building Blocks to High Level Functionality*, 26–33. Stanford, California: AAAI Press.

Arzi-Gonczarowski, Z. 2004. Let M be an embodied mind. In Davis, D. N., ed., *Visions of Mind (Forthcoming)*. http://www2.dcs.hull.ac.uk/NEAT/dnd/visions/mind.html.

Brooks, R. A. 1999. *Cambrian Intelligence: The Early History of the New AI*. Cambridge, MA: MIT Press.

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

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

Devlin, K. 2001. The real reason why software engineers need math. *Communications of the ACM* 44(10):21–22.

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

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

Isaacs, A.; Dainntith, J.; and Martin, E., eds. 1996. *Concise Science Dictionary*. Oxford University Press, third edition.

Lipson, H.; Antonsson, E. K.; and Koza, J. R., eds. 2003. *Papers from the 2003* AAAI *Spring Symposium on Computational Synthesis: From Basic Building Blocks to High Level Functionality*. Stanford, California: AAAI Press.

Sloman, A. 2004. The cognition and affect project. http://www.cs.bham.ac.uk/research/cogaff.