AI Emotions: Will One Know Them When One Sees Them?

Zippora Arzi–Gonczarowski
Typographics Ltd, Jerusalem 96222, Israel

Final draft 2002

Abstract

A systems approach to intelligent agent construction is proposed, where one starts from basic reactive mechanisms and proceeds to upgrade and to improve the system step by step. Along the way behavioral phenomena emerge that one could perhaps classify as emotions when one sees them.

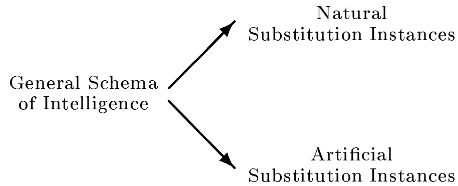
1 Circumscribing the Issue

If evolution gave us emotions - downright, abundantly, and painfully - then they should somehow be relevant to functional roles in the survival of the human species, and hence to its intelligence. However, this is not necessarily a reason to simply patch emotions onto artificial agents to make them intelligent, as this could be putting the cart before the horses. In this paper a systems approach to formal agent construction is proposed, in an attempt to figure out the literal identity of the cart and the horses in this metaphor.

The current discussion is about artificial intelligences that could possibly gain something by growing their own, innate, version of emotions. A few other domains of research, also related to emotions and to an artificial context, deal with different issues, such as: ($\imath$) Artificial agents that recognize symptoms of emotions in humans [19]. ($\imath\imath$) Artificial animation of emotional displays for purposes of entertainment, or social interaction with humans [11]. ($\imath\imath\imath$) Synthetic testbeds that could perhaps be designed to examine theories about human emotions [12]. These are important and interesting domains of research that could perhaps be classified as humanistic in nature: Even when the *anima* is generated inside an artificial agent, it is designed to convince humans, to interact with humans, to emulate humans, and so on. This study is about emotions in a general context of intelligence. From this perspective, humans are but an example, a substitution instance of a schema: tantalizingly successful, but not necessarily flawless, and not necessarily a unique option. Similarities and analogies between human and artificial agents would be by virtue of the same general schema that they instantiate (see figure), not because one imitates the other.

1.1 Nuts to Crack

($\imath$) It is hard to engineer anything that one cannot precisely define. There is no rigorous and general definition of emotions, in spite of millennia of attempts.



One simply knows emotions when one sees them. However, a *you know it when you see it* definition could perhaps apply to the recognition of something, but hardly for its generation.

(*ii*) Phenomena from one species are not necessarily suitable for another species, an artificial species in the present context. What one is actually looking for is some kind of *analog* of emotions for that other species, and one should be aware that the term 'emotions' is being used in a metaphorical sense. What is their literal sense for artificial agents? Will one know them when one sees them?

(*iii*) There is neurological evidence that emotions are tightly intertwined with other mind phenomena. Therefore they should not be patched as an addition to the minds of agents, but rather (if at all necessary) be integrated all over. For that, one first needs some analytic understanding of intelligence.

(*iiii*) It is not impossible that the functions of emotions could perhaps be played also by substitute mechanisms that do not necessarily look like emotions. It could be more reasonable to simply study and engineer intelligence. If emotions emerge, then perhaps one will know them when one sees them.

It is important to distinguish between, on one hand, cases where artificial agents just seem to have emotions because of their human users' tendencies to animism, and human users' natural needs for communication and understanding, and, on the other hand, innate artificial (analogs of) emotions.

It has been suggested, for example, that agents should be endowed with 'frustration'. Namely, if they keep failing in the pursuit of a goal, they should get 'frustrated', and, as a result, adjust their behavior as humans (sometimes) do: Try even harder, or divert to other methods, or divert to other goals, and so on. Before agents are burdened with a complex emotion just because humans tend to have it, one needs to be convinced that, for example, the same mechanism could not be activated when a counter of failures gets incremented beyond a certain threshold (is there a radically different way to emulate that in software?) In that case, a human programmer might perhaps name a relevant procedure call '`get_frustrated`', because this helps humans understand what is going on in the program by relating to their own experiences. For debugging or control purposes, the program could perhaps even output `'Hello world, I am frustrated'`. However, all that does not mean that the program is emotionally frustrated, or that it is emotional at all. We do have a creative programmer who metaphorizes. (The introspective programmer might have preferred to be endowed with a counter of failures that activates a diversion procedure, rather than go through the unpleasant visceral experience of being frustrated. Maybe human architecture came up with frustration just because it does not support registers and counters.) It has been shown that exaggerated abuse of 'emotional' outputs like the above could irritate users. The novelty of the 'emotional' message eventually wears out, and they want to turn the feature off [20]. As in their private lives, these users do not care for repetitive emotional displays without the real thing behind them (one knows the real thing when one sees it).

These reservations could indeed be negotiated, and some of them possibly refuted. However, even if one gets convinced that giving agents human EMOTIONS, *in toto*, is somehow the right thing to do, then the concept EMOTIONS is still overwhelming. Would one honestly know how to do that or where to start? So one needs to look round for a beginning.

1.2 The Treaded Track

One effective tradition of foundational scientific research has been to go back to first principles in order to grapple with an issue. Intelligence is the end. What are the first principles of intelligence? [1] says: *'a prerequisite for something to be intelligent is that it has some way of sensing the environment and then selecting and performing actions.'* Intelligence hence boils down to a sensible marriage between behavior and circumstances.

Agents' concerns are first of all about survival, and one may also add the pursuit of various goals. The behavior that is generated in the service of these concerns should be autonomously initiated from inside the relevant agent, or else the relevant intelligence should rather be attributed to the external entity that drives the agent. Behaviors are more than often conjured up as responses to stimuli in the environment, hence agents are provided with a sensing apparatus.

In the biological context, for example, evolution naturally selected sensory motor neural apparatuses that coupled embodiments of organisms with their ecological niches, yielding behavior designated as 'intelligent' because it happened to support endurance of the species. In the artificial context agents are typically constructed to serve a purpose, so that 'intelligent' behavior is goal-directed. However, survival is often a concern in that context as well: The setting of agents in external environments exposes them to hazards that could not always be expected. Material existences in real physical environments as well as virtual entities in 'cyber' environments are in jeopardy. They can be injured and incapacitated. In dynamic environments some of the protective measures should be typically reactive: agents should be able to sense danger as it comes and to react, often urgently, in an appropriate manner to safeguard their existence. In both natural and artificial contexts, sensations and reactions should be tightly coupled, as they determine each other: Suitable reactions are conjured up by discriminating sensations that are, in turn, tailored for the forms of behavior that are afforded by the agent.

It has long been accepted that forms of natural intelligence, including sublime ones, are results of (combinations of) improvements and upgrades applied, by natural selection, on top of basic reactive intelligence. A similar modeling strategy is proposed here (definitely not for the first time) for an orderly and systematic approach to the construction of artificial agents. However, it is *not* suggested that AI should repeat the disorderly 'method' of natural evolution, just the general principle of grading is being advocated.

Natural selection upgraded intelligent agents from simple reactive organisms in a random and cluttered way, patch over patch. The results are living proof that scruffy design works, marvelously. One has to admit, though, that development was slow, the structure generally hard to decypher, and failures are often tragic and extremely difficult to debug. With some inspiration from these formidable natural examples, one might try an orderly approach when given a chance to consciously and systematically design artificial intelligences, keeping a neater record of that which goes on, and sorting out the cart after the horses.

2 Upgrading Reactive Behavior

Starting from (a general model of) basic reactive mechanisms, below are a few suggestions for upgrades that could be possibly structured on top of that. That these capabilities could be significant to survival and to intelligence is *by no means a novel idea*. The purpose here is to highlight their essence as upgrades and improvements that could be naturally structured on top of basic reactive mechanisms.

A basic notion that is applied here is *action tendency*, which is an internal incitement to perform a certain *behavior*. Action tendencies may become consummated actions, or not, depending on a variety of reasons. ([15] defines the core of an emotion as the readiness to act in a certain way.) If the descriptions below somehow remind readers of emotions, then one will know them when one sees them.

2.1 Upgrades

Excitation and Perseverance. Action tendencies might be prevented from becoming consummated actions by various obstacles. That could put agents' concerns at risk. There is hence reason to upgrade agents with perseverance: Behavior excitation should be vigorous and persist through thick and thin. Example: if something blocks the emergency exit during an emergency, then an agent should forcefully push the obstacle in the course of its flight.

In physically embodied agents, that materially interact with their environments, implementations of that - mechanical, thermodynamic, chemical, and so on, could not be instantly turned on and off. They would take time to build up and to accelerate, also to deccelerate and to dissolve when no longer needed. In humans, for example, reactions to alarms are precipitated, and it does take a while to relax back to normal after a stimulating experience. That could often be shown physiologically (e.g. by measuring heart rate). We even use metaphors from physical and technological processes to describe human excitation: heat, stir, fire, steam, shake... vs. let off steam, cool down... For artificially engineered agents, tension, pressure, stress, or strain, could turn out to be literal. If a spring gets deflected beyond a safety margin, approaching its proportional - or elastic - limit, then it would be sensible to activate a diversion procedure. That's innate.

Learning and adaptation. Following cases where behaviors have been, or should have been, conjured up, and agents (hopefully) survive, lessons could perhaps be learned. There is hence reason to upgrade these systems with contingent mechanisms that update relevant discriminations, remodel behaviors, or adjust degrees of perseverance and their safety margins. As a result, agents that feature that upgrade may develop different behavior patterns after having been exposed to different experiences at run-time, even if they started off from the assembly line with identical designs. Example: In an emergency, agent x did not push the package that blocked the emergency exit hard enough, stayed in, and suffered serious damage that needed repair. Next time agent x might tune itself to push harder. Meanwhile, agent y went through a different experience: A package that surrendered to an obstinate push, contained a priceless equipment that broke. Next time agent y might learn to discriminate, and to attend to, 'fragile' icons.

Integrating and handling conflicts. If more than a single reaction is conjured up simultaneously (e.g. as a result of multiple stimuli) and not all actions can be technically performed together, confusion arises, triggering disordered responses. Example: an emergency signal prompts the agent to urgently push the package that blocks its passage, while another stimulus (e.g. a 'fragile' icon) incites mechanisms that impede the agent from touching the package. Agents may not feature higher-level reasoning (an upgrade that will be discussed below) to decide what to do, and, in any case, urgency often debars the option of a leisurely consultation with higher-level reasoning.

Upgraded agents could feature higher level impulses, that are structured on top of the basic reactions, and are designed to arbitrate between multiple, possibly conflicting, behaviors. This may consist, for example, of a mechanism of automatic prioritization and selection (e.g. survival is most important, so it overrides everything). A more complex possibility would be to automatically substitute or integrate essential elements from a few behaviors into one coherent behavior that compromises a little, but takes care of (almost) everything (e.g. take a minute to move the fragile obstacle very carefully, then speed up as much as possible to make up for the lost minute).

Solutions to conflicts would often involve a (partial) suppression of the primary reactions (e.g. to push the package vigorously), and applying this control would require energy and time in the presence of perseverance. That perseverance, which has just been argued for rationally, should hence be controllable in some situations. One programming solution could be to condition the activation of perseverant behaviors: `perform(x)` <u>`unless`</u>`...` However, additional observations, that are required to watch out for a long list of reservations and conditions, could come at the expense of urgency. The list could turn out to be too long, or, even worse, unpredictable. It is more practical to let the exceptions take care of themselves and let them apply whatever is needed to counteract other perseverant behaviors that come their way.

Upon tergiversations of perseverant action tendencies, there will always be a period of dissolution as the primary impulses persist while they are slowly fading out, and energy is being invested in containing that arduous process. An agent that prioritizes or integrates, may be in control of its behavior, but that control takes energy resources that need to be accounted for over a period of time. The machinary may seem to be inert, but inside it various mechanisms may be active at their full capacity just to cancel one another, draining energy resources, causing internal wear and tear. Locally, this may not seem rational, but it could be the result of a globally sensible design.

General rules for tuning degrees of controllable perseverance that are *just right* and take care of all the considerations above, and of all possible run time scenarios, seem to be very hard to formulate. If that is indeed the case, then locally irrational turbulances and behaviors will sometimes emerge.

Memory and Anticipation. More beneficial than getting into and out of trouble, would be to avoid compromising situations in advance. Similarly, an agent could perhaps follow trails to incur advantageous situations that serve its concerns. For such purposes, agents could perhaps be upgraded to conceive of plausible future developments of current situations, and to remember past experiences. Reactions to those internal virtual images should then yield a more intelligent behavior by these agents. Example: on the basis of conceived, or memorized, images of emergency cases where emergency exits are blocked, these agents would always keep the emergency exit clear and never put anything in front of it. Even on untroubled days, they would react with telling effect whenever anybody else does anything to block an emergency exit.

Cognition. Memory and Anticipation, as suggested above, require an internal representational apparatus. Agents could be upgraded with the possibility to encode sensations internally, to reference, access, organize, and process the internal representation in various ways. Example: plan emergency exits during buildings design. Upon examination of a design, an agent following the analogy between the blueprint and its environmental realization, will be able to judge the plan by *reacting to a conceived image of an imagined emergency in an imaginary building.* (In [21] that phenomenon is designated 'as if' emotions.)

Reasoning. Given internal representations and time, behavior could be based on meticulous high-level reasoning that consider and carefully weigh all available aspects and consequences of a situation: Agents may not just run to an emergency exit, but carefully reason about the optimal evacuation trajectory, analyze the pros and cons of compatible options, define and study fallback options, perhaps generate an entire taxonomy of those that is based on added discriminations and calculated values of various parameters, tradeoffs, and related formulae. If, and when, this is possible, results are expected to yield remarkably reliable behaviors. It follows that agents should *persevere in their attempts to apply that kind of intelligence*, and to compete with other behaviors for the necessary resources. In the human context, for example, agents (like absent-minded scientists) could get deeply absorbed in calculations to a point where they may not attend to meaningful events in their environments.

Social behavior and communication. A social environment consists of other behaving agents. They are different from other environmental elements in that they also feature a behavior that is based on contingent, autonomous, reactions. Just imagine, for instance, that an emergency exit happens to be blocked not by an inanimate object, but rather by a fellow agent, featuring a mind of its own, who wouldn't move. Agents could be upgraded to make discriminations and predictions about the behavior of other agents, and to benefit from that. When agents are also upgraded with anticipation, for example, subject agents could perhaps predict that emergency situations change object agents' usual behavior. Example: 'In an emergency, another agent might not notice the 'fragile' icon'. If object agents also do the same about the subject agent, additional iterations of this cycle are possible. Better upgraded agents should perhaps be one (or more) iteration ahead of the less upgraded agents. Examples: 'This is an emergency, so no one expects me to notice the 'fragile' icon', or: 'This was an emergency, so the agent was counting that no one should expect it to notice the 'fragile' icon.'

Self perception. The latter kind of upgrade could be applied to the subject agent itself, yielding a more sophisticated kind of higher order control: It could perhaps behave also on the basis of discriminations and predictions about itself and its own mechanisms. Example: 'In an emergency, I am programmed to pay attention to nothing, so I should always keep the emergency exit clear'. In fact, by virtue of each agent being a disjoint entity with boundaries, it is naturally expected to enjoy a better grasp of its own mechanisms than those of others. Example: 'My emergency mechanisms are activated'. In fact, [17] argued that emotions are perceptions of physiological reactions by the body, and for [13] *emotions* are reactions, distinguished from *feelings*, which are their conscious perceptions.

The above engineering perspective of improvements and upgrades of basic reactive behavior (there are indeed more types and variants) is essentially about management, maintenance, and amelioration of a large household of adamant action tendencies. That economy requires more action tendencies, and higher order ones, so that the system's behavior, that is finally and actually generated, should be sensible. A significant design principle is that one is not allowed to deny the legitimacy, or get rid, of the lower level action tendencies. One is only allowed to toy with smarter, and more adamant, controllers, arbitrators, diverters, negotiators, reasoners, and so on.

2.2 Derivative Upgrades

Improvements and upgrades as above do not necessarily take place together. They are likely to co-occur if the same, or similar, system modifications avail more than one upgrade. The significance and the value of a system feature obviously increases with the number and the extent of the capabilities that it supports. Quite a few of the features that are suggested above depend on the possibility to represent environments internally and to toy with that representation, to the extent that it has become a common mistake to substitute that capability for intelligence itself.

When a few upgrades do happen together, and are intertwined in one agent, intelligent behavior would hopefully excel. However, surprising, possibly uncalled for, happenstances could also occur:

(*ı*) There is no apriori guarantee that the superposition of various improvements should not give rise to undesired effects, that are hardly related to agents' concerns. In the human context, for example, behavioral phenomena such as religious practices, scientific curiosity, artistic expression, abstract mathematical reasoning, as well as obsessions with those, are but a few examples of prevalent behaviors that are not clearly related to evolutionary survival, but they happened. Whenever design considerations require a subset of intelligent upgrades, then one should perhaps also live with the results of the superposition.

(*ıı*) Reactions could be manipulated by other intelligences, be abused against the agent. Some ancient human martial arts, for example, are based on manipulations of motor reactions in the opponent. (The possibility of manipulation should remind us that *autonomy* is indeed a philosophically difficult concept.)

(*ııı*) Existing capabilities could be rather easily interconnected in novel ways. Evolution theorists use the term *exaptations* [16] to refer to minor changes that make use of already existing capabilities to create new behaviors. Exaptations in the context of the current discussion could be considered:

(*ı*) The capability to hard wire a reaction to a stimulus is basically about direct physical reactions. When agents are upgraded with internal capabilities, such as memory or reasoning, then these additions could also get hard wired as reactions, and that would be an exaptation of the basic capability to react.

Memories of similar past experiences, for example, could be conjured up in a reactive, non-deliberate way.

(*ii*) The basic requirement that agents should leave the assembly line ('be born') with certain innate reactive mechanisms, could possibly exapt to other innate motivations. In the human context, for example, agents who react only when necessary, or only in pursuit of predefined goals, are considered undermotivated, dull, maybe even depressed. To make the most out of the capabilities of agents, they could be designed with innate motivations to exhaust their capabilities also when they are not specifically called for. As motivations for free exploration, for example, become perseverant, they could eventually compete with other action tendencies. Human history abounds with people who risked their lives to explore.

(*iii*) The mechanism that arbitrates between conflicting reactions could exapt to arbitrate in non urgent situations as well. It could happen, for example, that higher level reasoning mechanisms cannot find a solution to a problem (some problems are undecidable or intractable, for instance). The impulsive arbitration mechanism could help, either by simply making the decision instead (in humans this is called decision by gut feelings), or by pruning some of the options to make the rational decision more tractable (in computational complexity theory this is modeled by an *oracle*). This is perhaps allusive of Minsky's argument that *Emotions are other ways to think* [18]. There is neurological evidence that pruning options is the role of emotional modules which participate in rational decision making [13].

Higher order upgrades, like self perception, seem to provide most of the surprises. [14] says: *'Formal mathematics is at most 5000 years or so old... not long enough for our brains to undergo any but the most minor changes. Thus, the mental processes we use to do mathematics must have been acquired and in use long before the Sumerians introduced abstract numbers... The new twist required in order to do mathematics was to bring those capacities together and use them to reason not about the physical and social world for which they initially developed through natural selection, but rather a purely abstract world of the mind's own creation.* But agents could also get derailed as they perceive themselves, and 'perturbant' emotions as defined in [21] are an example.

3 Mathematical Modeling

The domain under consideration should be especially suitable for description using a mathematical modeling language. The principle of mathematical method is to start from fundamental concepts as primitive terms, and assert certain simple propositions (postulates, axioms) about them. Further terms are then introduced in an orderly manner, using the primitive terms. Theorems express properties of these new terms that are assured by this mode of orderly generation, and deductive reasoning is used to obtain their properties from the postulates. As a result, starting from premises which are intuitively convincing and obvious, and following a long series of simple steps, each convincing by itself, one obtains complex constructs, and truth for propositions, that would be far from obvious had they been asserted at the outset. The typical paradigm is the system of natural numbers. A suitable selection of properties constitute the five postulates of Peano, capturing the pre-theoretical essence of the natural numbers as concepts that are used for counting discreet objects. Orderly extensions of the natural numbers provide the integers, then the rational numbers, then the real numbers.

Hilbert coined the term *The Genetic Method* for that, which is suggestive for the current discussion: Natural intelligent systems started evolving from the earliest nerve cell that was probably a combined receptor (receiving environmental stimuli) and motor unit (producing muscle or gland response). With biological systems as role models, intelligent systems should be designed by starting from primitive terms that model an abstraction of that, and orderly structured extensions should then be introduced to model intelligent upgrades, using deduction to obtain their properties.

A bimodality of mathematical modeling is that, on one hand, one should treat the concepts *as if* they were meaningless, to ensure that all assumptions be stated explicitly as postulates, and that no hidden properties of the primitive terms should enter from their pre-theoretical, commonsense, intuitions. This is significant if the theory is going to serve engineering purposes. On the other hand, one needs to invariably refer to the domain that is being modeled, making certain that the theory that emerges *validly* describes the phenomena that gave rise to the formalization.

Modularization is often a means to master complex systems. Its danger, losing the integrative essence of systems, is more than just hypothetical in view of the fragmentation of AI research. Mathematical modeling enables ramified modularizations that do not lose the integrative essence of systems, because all components are based on the same premises.

ISAAC, an Integrated Schema for Affective Artificial Cognition, [10, 9, 2, 4, 3, 6, 5, 7, 8] is a mathematical formalism that follows these guidelines. It boots a model of agents' 'minds' from (a formalization of) reaction driven perceptions, and then models various upgrades on top of that, using a rigorous mathematical framework. A relatively small number of mathematical concepts and constructs avail quite a few upgrades, allowing various capabilities to exapt and to co-occur in one system. It is beyond the scope of this paper to describe ISAAC, and readers are referred to the citations above.

4 Human Emotions Revisited

A formal structure of general intelligence is likely to provide us with a formal ontology, designating and classifying behaviors, action tendencies, and related concepts. (Some of the behavioral ontology that is generated by ISAAC's formalism was described in [6].) Such an ontology could be then applied to the human substitution instance, possibly yielding an ontology for human behavior that is general, rigorous, and, hopefully, not too biased by the fact that the researchers themselves are humans with subjective human experiences. (Such biases are legitimate, provided that a clear distinction is made between the general schema, where they would be over deterministic, and the specific substitution instance where they occur.) If that formal ontology happens to coincide with theories of human behavior, if it yields descriptions of phenomena that one could classify as emotional (because one knows them when one sees them) then this would be tying the ends of our pre-theoretical intuition, that emotions are an essential part of intelligence. Following the principles of mathematical modeling as discussed in section 3, that would be a reinforcing feedback that the metaphorical cart and horses (from the introduction to this paper) are being arranged in their correct order on the track.

References

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

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

[3] 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.

[4] 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.

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

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

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

[8] Z. Arzi-Gonczarowski. Self, empath, manipulativity: Mathematical connections between higher order perception, emotions, and social cognition. In D. Cañamero, editor, *Papers from the 2001 AAAI Fall Symposium: Emotional and Intelligent II – The Tangled Knot of Social Cognition*, pages 9–14, Cape Cod, North Falmouth, Massachusetts, November 2001.

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

[10] 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.

[11] R. Aylett and L. Cañamero, editors. *Proceedings of the AISB'02 Symposium on Animating experessive Characters for Social Interaction*. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, UK, 2002. http://homepages.feis.herts.ac.uk/ comqlc/aecsi02.

[12] L. Cañamero. Emotions and adaptation in autonomous agents: A design perspective. *Cybernetics and Systems*, 32(5):507–529, July-August 2001.

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

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

[15] N. H. Frijda. *The Emotions.* Cambridge University Press, Cambridge, 1986.

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

[17] W. James. What is an emotion? *Mind*, 9:188–205, 1884.

[18] M. Minsky. The emotion machine. http://www.media.mit.edu/people/minsky, 2002.

[19] R.W. Picard. *Affective Computing.* The MIT Press, 1997.

[20] R.W. Picard and J. Klein. Computers that recognize and respond to user emotion. *Interacting with Computers*, 2002. forthcoming.

[21] A. Sloman. The cognition and affect project. http://www.cs.bham.ac.uk/research/cogaff/0-INDEX.html, 2002.