Using Surprise to Create Products that get the Attention of other Agents

Luís Macedo[1,2], Amílcar Cardoso[2]

[1]Instituto Superior de Engenharia de Coimbra
Quinta da Nora
3031-601 Coimbra - Portugal
lmacedo@isec.pt

[2]CISUC - Centro de Informática e Sistemas da Universidade de Coimbra - Dept. Engª Informática da Universidade de Coimbra
Pinhal de Marrocos
3030 Coimbra - Portugal
{macedo, amilcar}@dei.uc.pt

Abstract

This paper addresses an aspect of social environments comprising a series of processes that begin with an artificial agent (the "author-agent") producing surprising products (objects, events, etc.), continue with other artificial agents (the "jury-agents") appraising those products with respect to surprise (manifested, for instance, by the focus of their attention on those products), and end with a possible update of the emotional state of the "author-agent" by the elicitation of emotions such as happiness and pride, or sadness (depending on whether or not those products get the attention of those "jury-agents"). We describe a model of surprise that is mainly rooted in the cognitive-psychoevolutionary model of surprise proposed by the research group of the University of Bielefeld (Meyer, Reinsenzein, Schutzwohl, etc.) and also in the ideas of Ortony and Partridge. We present an experimental test about the activity related to the creation and evaluation of surprising objects.

Introduction

Considered by many authors as a biologically fundamental emotion (e.g.: Ekman 1992; Izard 1977), surprise may play an important role in cognitive activities, especially in attention focusing, learning (Izard 1977; Meyer, Reisenzein, and Schutzwohl 1997; Ortony and Partridge 1987; Reisenzein 2000) and creativity (Boden 1995; Macedo and Cardoso 2001b) (note however that some authors, like Ortony, Clore and Collins (1988), do not consider surprise an emotion).

Experiments performed by the German research group of the University of Bielefeld (e.g.: Meyer, Reisenzein, and Schutzwohl 1997) provide evidence indicating that surprise-eliciting events initiate a series of mental processes that begin with the appraisal of a cognized event as exceeding some threshold value of unexpectedness, continue with the interruption of ongoing information processing and the reallocation of processing resources to the unexpected event, and culminate with an analysis and evaluation of that event plus immediate reactions to it and/or belief revision. According to these authors, surprise has two main functions, informational and motivational, that together promote both immediate adaptive actions to the surprising event and the prediction, control and effective dealings with future occurrences of the event.

Although agreeing that surprisingness and expectation failure are closely related concepts, Ortony and Partridge (1987) argue that there are many situations in which one is surprised by something one didn't expect without having to expect something else. Ortony and Partridge's view of surprise shares aspects with the model proposed by the research group of the University of Bielefeld, especially in that both assume that unexpected events elicit surprise. The same is also true for Peters' (1998) computational model of surprise, implemented in a computer vision system, that focuses on the detection of unexpected movements. Other authors (e.g.: Breazeal and Velásquez 1998) have also modeled surprise in their works.

One of the features of humans is the need to create things (objects or events) that get the attention of other humans. One of the ways of achieving this is by creating surprising things. Besides surprise, other variables such as novelty also influence the attention of humans (Izard 1977). The role of these two variables, surprise and novelty, is particularly important in creative activity. Originality, which may be defined as unexpected novelty, i.e., as surprising novelty, is consensually accepted as one of the distinguishing characteristics of creative products/ideas (Boden 1995, Macedo et al. 1998), in addition to appropriateness (i.e., usefulness, aesthetic value, rightness, etc.). Besides surprise, other emotions seem to be involved in creative activity (see Picard 1997). At least, emotions such as happiness and pride seem to be rewards of producing creative products. As implied, to some extent, by Ortony, Clore, and Collins' cognitive theory of emotions (1988), these seem to be elicited not only when the creative product satisfies the personal needs of the author, but also when the author perceives that other agents consider that product as a creative one, which may be manifested by the expression of surprise and the focus of attention on that product. In opposition, sadness seems to be elicited when those products do not get the attention of other agents.

In this paper we focus on the activity of artificial agents concerning the production of objects that, causing surprise to other artificial agents, get their attention, and consequently influence the emotional state of the "author-agent" itself.

The following section introduces the multi-agent environment. Subsequently, we present an overview of the overall architecture of an agent integrating a model of surprise. Afterwards, we present the reasoning process of an "author-agent". Then, we describe an experimental test. Finally, we conclude by discussing the limitations of the model and presenting future work.

A Multi-agent Environment

We have developed a multi-agent environment in which there are two kinds of artificial agents interacting in a simple way: the "author-agents" whose main function is to create things (objects, events), and the "jury-agents" whose goal is to explore the environment, analyzing, studding and evaluating it. Additionally, there are objects located at specific positions (see Figure 1). In this paper, these objects are confined to buildings. Each object comprises three distinct, fundamental components: *structure*, *function* and *behavior* (Goel 1992). For the sake of simplicity, the *structure*, the visible part of the object, is restricted to the shape of the object (e.g., triangular, rectangular, etc.). The *function* of the object describes its role in the environment (e.g., house, church, hotel, etc.). The *behavior* of the object is related to its activity (actions and reactions) in response to particular features of external or internal stimuli (e.g., static, mobile).

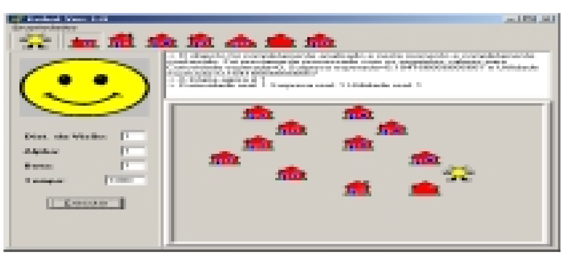

Figure 1 - Example of an environment.

We have been using this environment to model an aspect of social environments comprising a series of processes that begin with an artificial agent (the "author-agent") producing surprising products (in this case buildings), proceed with the possible elicitation of surprise by those products in other artificial agents (the "jury-agents"), manifested mainly by the focus of their attention on those products and subsequent analysis and/or evaluation of them, and end with a subsequent reversal influence of the behavior of those "jury-agents" on the emotional state of the "author-agent".

Overview of the Architecture of an Agent

A possible architecture for a social agent that takes surprise into account in its reasoning/decision-making is depicted at a high level in Figure 2. It comprises the following modules (explained in more detail in the subsections below): *sensors/perception*; *memory*; *emotions, drives and other motivations*; and *reasoning/decision-making*. These last two modules are provided with information from the world obtained through *sensors/perception* and also recorded in *memory*. Then, the *reasoning/decision-making* module computes the current state of the world (external and internal). Afterwards, Probability Theory (Shafer and Pearl 1990) is applied to predict possible future states of the world for the available actions (internal or external), and an Utility Function (Shafer and Pearl 1990; Russel and Norvig 1995) (which makes use, for instance, of the intensity of the generated emotions) is applied to each one of those states of the world. Finally, the action that maximizes that function is selected.

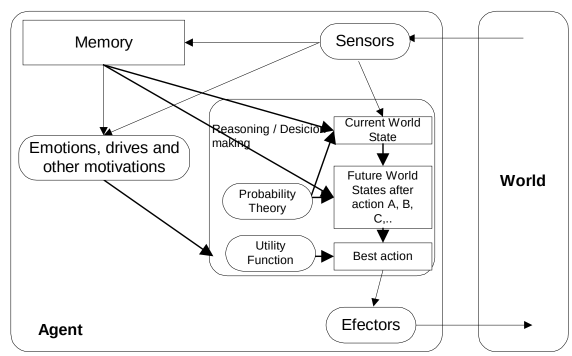

Figure 2 - Agent's architecture. The ovals represent processing modules while the rectangles represent information modules.

Sensors/Perception

The information related to the *structure*, the *function*, the *behavior* and the distance of the objects is collected from the environment through simulated sensors. There is an user definable parameter for the range of the visual field. Objects out of that range are not visible by the agent. The *function* of the objects is not accessible (i.e., cannot be inferred from visual information) unless the agent is at the same place as the object.

Action

The agent has two main activities: the creation of products and its addition to the environment, and the exploration of the environment (see Macedo and Cardoso 2001a, 2001b for more details about these activities). Actions associated with these activities include both the addition of pieces to the product that is under construction and the movements to certain locations of the environment.

Goals

The ultimate goals of the agent, namely the exploration of the environment and the creation and addition of products to the environment, result from the achievement of another goal: the maximization of positive feelings and the minimization of negative ones. Thus, we are considering emotions as action-goals (Reisenzein 1996).

Memory

The agent's knowledge base is of an episodic kind: each object is stored as an individual case in the episodic memory and associated with a number that expresses its absolute frequency (see Figure 3). In addition to this case-base, the agent also has a map of the environment in memory where it stores the location of the objects.

Field \ Case	C_1	C_2	C_3	C_4
Structure				
Function	House	House	Church	Hotel
Behavior	Static	Static	Static	Static
Abs. Freq.	50	40	5	5

Figure 3 - Example of the episodic memory of an agent after exploring an environment.

Emotions, Drives and other Motivations

In this paper, the emotional makeup of the agent is confined to surprise and also to simplistic forms of happiness, sadness and pride. Nonetheless, part of our ongoing work involves the inclusion in this module of additional emotions, drives and yet another motivations.

Following Ortony and Partridge (1987), the agent is almost continuously presented with an *input proposition*, which in the case of the environment described above corresponds to some information (visual or not) of a building (for instance, "a house with squared windows"). We also distinguish between *deducible* and *non-deducible*, *active* and *passive*, *immutable* and *typical* propositions. The immutability of a proposition can be extracted from the absolute frequency values associated with the cases (see Figure 3 above). For instance, the proposition "houses have squared facades" is immutable (since all the houses in memory have squared facades), whereas "houses have squared windows" is a typical proposition with an immutability (probability) value of .55 (as implied by Ortony and Partridge's model, in our model immutability is a continuous variable).

We share Ortony and Partridge's view that surprise may result from three situations: (i) *active expectation failure*, resulting from a conflict or inconsistency between the *input proposition* and the *active prediction* or *expectation* (i.e., propositions explicitly represented in memory); (ii) *passive expectation failure* (or *assumption failure*), resulting from a conflict or inconsistency between the *input proposition* and what the agent implicitly knows or believes (*passive expectations or assumptions*), i.e., propositions that are not explicitly represented but that may be easily inferred; (iii) *unanticipated incongruities* or deviations from norms, resulting from a conflict or inconsistency between the *input proposition* (in this case a *practically non-deducible proposition*) and what, after the fact, may be judged to be normal or usual. Notice that, in this case, there are no (*passive* or *active*) *expectations* with which the *input proposition* could conflict (at least prior to the unexpected fact).

When the agent sees the structure of a building it computes expectations (*deducible*, *active expectations*) for its *function* (e.g., "it is a hotel with 45% of probability", etc.). If, after visiting that building, the agent finds out that it is a post office, it would be surprised, because its *active expectations* conflict with the *input proposition* (note that, in our model, belief conflicts may be partial as well as total). This is thus an example of the first source of surprise distinguished by Ortony and Partridge. In contrast, when the agent sees a building with a window (or roof, etc.) of a particular shape (for instance, circular), it is able to infer particular probabilities for each possible shape (a rectangular shape with, for instance, 45% probability, a squared shape with 67%, etc), although it may not have made an *active prediction* for them. These are examples of *deducible*, *passive expectations*: although not made before the agent perceived the building, it could easily infer expectations for the shape of the window after it was perceived. This case is therefore an example of the second source of surprise because the *input proposition* "has a circular window" conflicts with the agent's *passive expectations*. Finally, when the agent sees a building with no facade, it has neither an *active* nor a *passive expectation* available, because there are no buildings with no facade in its memory and therefore the agent could not predict that. Thus, "the house has no facade" is an example of a *non-deducible proposition*. This is an example of the third source of surprise: there is a conflict between the *input proposition* "the house has no facade" and what after the fact is judged to be normal or usual ("buildings have a facade").

We have implemented a computational model of surprise that is an adaptation (although with some simplifications) of the University of Bielefeld's model (e.g.: Meyer, Reisenzein, and Schutzwohl 1997) and in which the following four mental processes elicited by surprising events are present: (i) the appraisal of a cognized event as exceeding some threshold value of unexpectedness; (ii) interruption of ongoing information processing and reallocation of processing resources to the investigation of the unexpected event; (iii) analysis/evaluation of that event; (iv) possibly, immediate reactions to that event and/or updating or revision of the "old" schemas or beliefs. The suggestions by Ortony and Partridge are mainly concerned with the first of these steps, and are compatible with the Bielefeld model (see Reisenzein 2000). Accordingly, we drew on these assumptions for the implementation of the appraisal of unexpectedness and the computation of the intensity of surprise.

Given an *input proposition*, the *surprise generation module* outputs the intensity of the elicited surprise, taking into account the memory of the agent. A correspondent facial expression is also produced (Ekman 1992). There is experimental evidence supporting that the intensity of felt surprise increases monotonically, and is closely correlated with the degree of unexpectedness (see Reisenzein 2000, for a review of these experiments). This suggests that unexpectedness is the proximate cognitive cause of the surprise experience. On the basis of this evidence, we propose that the surprise felt by an agent *Agt* elicited by an object Obj_k is proportional to the degree of unexpectedness of Obj_k, considering the set of objects present in the memory of the agent. According to Probability Theory (e.g.: Shafer and Pearl 1990), the improbability of X, denoted by 1-P(X), defines the degree of not expecting X, and the intensity of surprise can, for simplicity, be equated with unexpectedness:

$$SURPRISE(Agt,Obj_k)=$$
$$DegreeOfUnexpectedness(Obj_k,Agt(Mem))=1-P(Obj_k)$$

Although other probabilistic methods might be used to compute P(X), in the case of objects comprising several components we propose to compute the probability of the whole object Obj_k as the mean of the conditional probabilities of their n constituent parts, which are individually computed using Bayes's equation (Shafer and Pearl 1990) (note that each one of those conditional probabilities individually gives the degree of unexpectedness of a specific piece of the object, given as evidence the rest of the object):

$$P(Obj_k)=\frac{\sum_{l=1}^{n} P(Obj_k^l \mid Obj_k^1,Obj_k^2,...,Obj_k^{l-1},Obj_k^{l+1},...,Obj_k^n)}{n}$$

Note that we have previously performed experiments that provide evidence indicating that the intensity values of surprise rated by an artificial agent with this model of surprise almost match the ones rated by humans under similar circumstances (see Macedo and Cardoso 2001c).

Happiness and pride are elicited when the agent perceives that the objects produced by it elicit surprise on and get the attention of the other agents present in the environment. In contrast, sadness is elicited when those objects do not elicit surprise nor get the attention of those other agents. The intensity of happiness and pride of an "author-agent" are, for simplicity, proportional to the number of agents that focused their attention on its products, while the intensity of sadness is inversely proportional.

Reasoning/Decision-making

The reasoning/decision-making module of the agent receives the information from the external world and outputs the action that has been selected for execution. This module comprises several subprocesses described as follows.

Computation of the current world state. Taking the information of the world provided by the sensors (which may be incomplete) as the input, the current state of the world (e.g., the agent's current position, the position of the objects, the shape of the objects, etc.) is computed.

Computation of future world states. Taking the current state of the world, Probability Theory (Russel and Norvig 1995; Shafer and Pearl 1990) and the memory-stored information as input, possible future world states and respective probabilities are computed for the actions that the agent can perform. These actions may be of two kinds: movements to certain locations in the environment (*MoveTo*(*Obj1*), *MoveTo*(*Obj2*), *MoveTo*(*LocationXY*), etc.) or addition of pieces to the product that is currently under construction (*AddPiece(X)*, *AddPiece(Y)*, etc.). For the former kind of actions, the resulting new world states include not only the new position of the agent, but also the information (e.g., relative position, shape, etc.) of the near objects, provided in that new world state. Instead, for the latter kind of actions, the new world states comprise the imaged or seen products (possibly partially constructed) resulting from the additions of pieces that have just been performed.

Usually, an action *A* may lead to one of a set of possible world states W_1, W_2, ..., W_n (it is not possible to know with complete certainty to which one, but it is possible to assign probabilities to them). This is described by what is called within Utility Theory as a Lottery (Russel and Norvig 1995; Shafer and Pearl 1990), which is represented by a list of elements, each one comprising a possible resulting state of the world and its associated probability:

$$Lottery(A)=\left[p_1,W_1;p_2,W_2;...;p_n,W_n\right]$$

where p_i is the probability of the i^{th} possible resulting world state W_i of the action *A*, and $\sum_i p_i=1$.

Selection of the "best" action. From those available actions, a single one (presumably the best one) is selected – the one with the highest Utility Value. These Utility Values result from the application of the Utility Theory as follows. For each action, the following Expected Utility Function, denoted by *EU*, is applied to its Lottery:

$$EU(p_1,W_1;p_2,W_2;...;p_n,W_n)=\sum_i p_i \times U(W_i)$$

where $U(W_i)$ denotes the Utility Function of state W_i.

This Utility Function relies heavily on the anticipated intensity of surprise elicited by the future state of the world. Thus, the preferences of the agent are reflections of its anticipated surprise:

$$U(W_i)=U_{surprise}(W_i)=SURPRISE(Agt,W_i)$$

The Reasoning Process of an "Author-agent"

Although the subprocesses of the *reasoning/decision-making* module described above are common to all agents, the reasoning process of an "author-agent" is to some extent special. Actually, we have implemented a Case-Based computational creative process which is partially inspired by psychological models such as the ones proposed by Wallas, Guilford, De Bono's, etc. (see e.g.: Macedo et al. 1998). This process involves the following four steps: problem acquisition plus knowledge assimilation, search for a product (or solution), proposal of a product, and verification of the proposed product (more details about this process and its application to music composition may be found in e.g.: Macedo et al. 1998). In this paper we consider solely the influence of surprise on that process. Therefore, we are not concerned with the achievement of creative products/ideas, but instead only with the production of surprising products, which may be a central aspect of creative production.

Roughly speaking, given a partially constructed product – the starting product - or even no product at all, an "author-agent" produces a surprising product reusing and combining *pieces* of products from past episodes (products present in memory). This is performed iteratively, i.e., piece by piece until the product becomes complete. In each iteration, the decision of what piece to add is not made by chance, but obeying to the Maximum Utility Principle (Russell and Norvig 1995) which is explained as follows. Given a set of possible actions that the agent might do, such as *Add piece X*, *Add piece Y*, the agent computes beforehand the states (in this case products or possibly partial products) resulting from each of those available actions, and then applies the Expected Utility Function described above. Considering that the goal is to produce surprising solutions, that function gives higher Utility Values to those actions leading to higher surprising products. A variety of products may be achieved repeating the process with different degrees of surprise, resulting a Divergent Production of products (see e.g.: Macedo 1998; Guilford, 1968). Figure 4 illustrates this process.

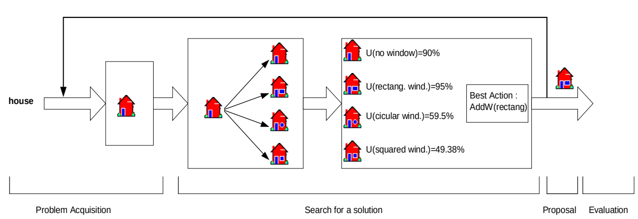

Figure 4 – Illustrative example of the decision-making process involved in the production of surprising products.

The set of products created by the "author-agent" through the process described above, and which up to now has been confined to the working memory of that "author-agent", may be now added to the environment and shown to other agents. Those agents, who are always moving from location to location, from object to object, exploring the environment, appraise the objects present in the environment with respect to surprise, and focus their attention and select visiting or analyzing those products that elicit more surprise. This is performed, as described above, applying the Expected Utility Function to the available actions of *MoveTo(Obj1)*, *MoveTo(Obj2)*, etc., and selecting the one with the highest Utility Value.

Depending on whether or not its products get the attention and are visited by other agents, the "author-agent" "feels" happiness and pride, or sadness, respectively.

Experiment

In the experiment that we performed, the "author-agent" was first asked to produce 8 buildings (denoted in Figure 5 by *A*, *B*, etc.). Its memory, the one presented in Figure 3, was always the same in all those 8 production processes. In contrast, the level of surprise required for those buildings changed (increasing from *A* to *H*): building *A* was produced with less surprise than building *B*, and so on. Subsequently, those buildings were provided to a set of 15 agents ("jury-agents"), selected after exploring part of the environment. Since those agents have different ages, they have different episodic memories, ranging from very small ones, with only 5 buildings (Agent 1), to large memories with 229 buildings (Agent 15), following an approximately linear distribution. For each agent, the intensity of surprise elicited by those objects was collected as well as the information of the object on which they focused attention. The main goal of this experiment is to take some conclusions about the relation between the level of surprise used in the production of objects and the intensity of surprise elicited by those products in the other agents, and also about the role of the amount of knowledge in the appraisal of those objects with respect to surprise.

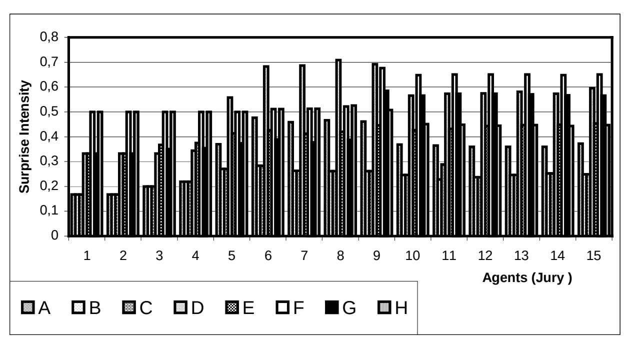

Figure 5 - Results of the experiment. For each one of the 15 "jury-agents", the surprise intensity elicited by buildings *A*, *B*, etc. (all produced by the same "author-agent") are represented by the columns from the left to the right, respectively.

According to the results of the experiment (Figure 5), we may conclude that, on average, the more the level of surprise used by an agent in the process of producing buildings, the more the surprise felt by the other agents, and the more the number of agents focusing its attention and visiting those buildings. Although it might be expected that agents with large episodic memories would "feel" less surprise than agents with small episodic memories, the experiment provides some evidence indicating that, on average, surprise is maintained although with a slight increase. Although this experiment does not give definite conclusions, we may explain this by the fact that agents with larger episodic memories may have a larger set of expectations for a given event, because of the larger knowledge they have, while agents with smaller episodic memories may have a smaller set of expectations for that event. Considering this experiment, building *F* is the one that deserves more attention by the other agents.

Conclusions and Future Work

Although the present model already addresses some important aspects of social agents, clearly much remains to be done. The environment is limited, the *Emotions, Drives and other Motivations* module of the agents is almost restricted to surprise, and the goals of the agents are confined to the exploration of the environment and creation of objects. We are currently working on the extension of the actual work in order to overcome these limitations: the environment is becoming more complex, containing other objects and events, in addition to buildings; additional emotions (fear, anger, etc.), drives and other motivations are being included in the *Emotions, Drives and other Motivations* module of the agents; the agents' goals are also being extended to comprise ordinary goals that are usually present in social environments.

Acknowledgments

We would like to thank Andrew Ortony and Rainer Reisenzein for their helpful discussions. This work was partially supported by POSI - Programa Operacional Sociedade de Informação of Portuguese Fundação para a Ciência e Tecnologia and European Union FEDER.

References

Boden, M. 1995. Creativity and unpredictability. *SEHR* 4(2).

Breazeal, C., and Velásquez, J. 1998. Toward Teaching a Robot "Infant" Using Emotive Communication Acts. In *Proceedings of SAB-98*.

Ekman, P. 1992. An Argument for Basic Emotions. In Stein, N. L., & Oatley, K. (Eds.), *Basic Emotions*, 169-200. Hove, UK: Lawrence Erlbaum.

Goel, A. 1992. Representation of Design Functions in Experience-Based Design. In: Brown, D., Walderon, M., Yosnikawa, H. (Eds.): *Intelligent Computer Aided Design*. Elsevier Science.

Guilford, J. 1968. *Intelligence, creativity and their educational implications*. San Diego, CA: Robert Knapp.

Izard, C. 1977. *Human Emotions*. New York: Plenum.

Macedo, L. 1998. A Model for Creative Problem Solving based on Divergent Production of Solutions. *Proceedings of the 13th European Conference on Artificial Intelligence*. New York: John Wiley & Sons.

Macedo, L., and Cardoso, A. 2001a. SC-EUNE – Surprise/Curiosity-based Exploration of Uncertain and Unknown Environments. In *Proceedings of the AISB'01 Symposium on Emotion, Cognition and Affective Computing*, 73-81. SSAISB.

Macedo, L., and Cardoso, A. 2001b. Creativity and Surprise. In *Proceedings of the AISB'01 Symposium on Creativity in Arts and Science*, 84-92. SSAISB.

Macedo, L., and Cardoso, A. 2001c. Modeling Forms of Surprise in an Artificial Agent. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, 588-593. Mahwah, NJ: Erlbaum.

Macedo, L., Pereira, F. C., Grilo, C., and Cardoso, A.1998. A Computational Model for Creative Planning. In Schmid, U., Krems, J., Wysotzki, F. (Eds.), *Mind Modelling: A Cognitive Science Approach to Reasoning, Learning and Discovery*. Berlin: Pabst Science Publishers.

Meyer, W., Reisenzein, R., and Schutzwohl, A. 1997. Towards a process analysis of emotions: The case of surprise. *Motivation and Emotion*, 21, 251-274.

Ortony, A., and Partridge, D. 1987. Surprisingness and Expectation Failure: What's the Difference?. *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, 106-108. Los Altos, CA: Morgan Kaufmann.

Ortony, A., Clore, G., and Collins, A. 1988. *The cognitive structure of emotions*. New York: Cambridge University Press.

Peters, M. 1998. Towards Artificial Forms of Intelligence, Creativity, and Surprise. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, 836-841. Mahwah, NJ: Erlbaum.

Picard, R. 1997. *Affective Computing*. Cambridge, MA: MIT Press.

Reisenzein, R. 2000. The subjective experience of surprise. In H. Bless & J. Forgas (Eds.), *The message within: The role of subjective experience in social cognition and behavior*. Philadelphia, PA: Psychology Press.

Reisenzein, R. (1996). Emotional Action Generation. In W. Battmann & S. Dutke (Eds.), *Processes of the molar regulation of behavior*. Lengerich: Pabst Science Publishers.

Russel, S., and Norvig, P. 1995. *Artificial Intelligence - A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall.

Shafer, G., and Pearl, J. eds. 1990. *Readings in Uncertain Reasoning*. Palo Alto, CA: Morgan Kaufmann.