An Architecture for Affective Behaviours Based on Conservation of Resources

Sabrina Campano, Etienne de Sevin, Vincent Corruble, and Nicolas Sabouret

Université Pierre et Marie Curie, Laboratoire d’Informatique de Paris 6
4, place Jussieu, 75005 Paris, France
{sabrina.campano,etienne.de-sevin,vincent.corruble,
nicolas.sabouret}@lip6.fr

Abstract. This paper presents a model for autonomous virtual agents that enables them to display affective behaviours. Our goal is to obtain believable behaviours, i.e. behaviours that are similar to those of human beings, for various simulation contexts in an urban environment. The proposed architecture is based on a principle of conservation and acquisition of resources.

Keywords: affect, emotion, virtual agent, simulation, behaviour

1 Introduction

Modelling believable behaviours is required to design human-like agents that can be used for credible simulations in domains such as security, urban planning, or video games. In this paper we present an architecture for autonomous virtual agents that enables them to display affective behaviours. The context is a realistic virtual city, with waiting lines as locations where conflicts can emerge between agents, and where danger like fire or riots can arise in the environment. Our objective is to define a model able to produce lifelike affective behaviours compatible with these situations.

Emotions have been considered as necessary components for lifelike virtual agents [1]. Inspired by psychological theories [2], some computational models rely on cognitive appraisal processes, in which a category of emotion is triggered by a specific context, and favours a set of cognitive strategies or behaviours [4, 6]. However, neither psychology nor computational science have come to an agreement on a basic set of emotions necessary and sufficient to cover the range of human behaviours: some claim for only two affective variables [7], others for six basic emotions [8], or even for twenty-two emotional variables [4], and their choice is justified by different criteria that all seem valid. This emotional parsing does not solve the issue of behaviour in computational models, since the same emotion is associated with multiple behaviours, and a behaviour can be associated with several emotions.

Considering these observations, and relying on the work of psychologist L.F. Barrett [11], we propose to view emotions as concepts independent from the
core architecture that generate behaviours. Our hypothesis is that emotions are categorizations that are useful for reasoning and communication purpose, but are not components at the origin of affective behaviour. Hence we aim at an affective architecture that should be able to generate behaviours describable with emotional lexicon by a human observer, without using emotional categories as components of the model. In this paper we propose a generic model for affective behaviours based on the theory of conservation of resources formulated by psychologist S.E. Hobfoll [13]. The central tenant of this theory is that humans try to protect their acquired resources, and seek to gain new ones.

After having considered related work, we present in this paper the core architecture of our model. The description of resources and how they fulfill an agent’s needs is explained, along with the selection process for resource-oriented behaviours. Finally we discuss the proposed model.

2 Related Work

Common sense lets one think that everyone knows what an emotion is, and that they are identifiable components of our brain system. “Fear”, “anger”, “joy”, “sadness” are words often used in our everyday vocabulary. However a close look at the litterature shows that emotions are all but natural kinds [11], and that there is currently no consensus on the number of existing emotions, neither on their role or consequences on cognition and behaviour [5, 14].

J. A. Russell [7] identifies only two types of core affect dimensions which are valence, i.e. how good or bad a feeling is, and arousal. He points out that any additional differentiation is based on contextual differences made upon various non-emotional processes. In a study on culture and categorization of emotions, Russell lists emotional words for which there is no equivalence from a language to another, revealing that emotion categories are culture specific, and that even the categories of fear and anger are not universal [9]. P. Ekman distinguishes among six basic emotions, grounded on the hypothesis of universal facial expressions, and on distinctive patterns of physiological changes during emotional episodes. These distinctions are still under debate, because even if autonomic specificity has solid support, it is difficult to match these patterns with definite emotional categories [3]. Besides, it is worthy of note that, according to R. W. Levenson, these studies do not prove the existence of emotions, but the existence of a correlation between an autonomic response and an emotional interpretation of this response by the subject.

Psychological theories of emotion are numerous and propose different emotion sets based on different valid criterias. S. S. Tomkins enumerates nine affects and three valences [16], adopting a functional approach of emotions, and Ortony, Clore and Collins (OCC model) account for twenty-six emotions using Lazarus’ evaluation theory [4]. The OCC model has been widely used in computational science. It aims at predicting which emotional category could be associated to a situation. However it is not suitable for simulating behaviours. In this model categories are considered as interpretations leading to cognitive strategies, or
what we may call reasoning. It is not possible to match a unique behaviour with
each emotion. For example fear and anger can both lead to agressive behaviours.
This results from the fact that Lazarus’ original model of cognitive appraisal is
centered around the question of how individuals interpret a situation to cope
with it, not at how universal emotion categories could trigger behaviours.

Besides the fact that emotions are culture specific [9], they are also individual
specific [10]. According to psychologist L.F. Barret, if no set of clearly defined
emotional patterns has been found, it is because emotions are concepts instead
of being distinct entities in our affective system [11]. Human beings experience
emotions the same manner as they experience colors, they use their knowledge
to label their perceptions with categories. Hence if emotions are concepts, it is
possible to parse our affective space with an infinite number of emotion sets.

From this conclusion, a question arises: what components an affective archi-
tecture generating behaviours labelled as “emotional” must have? Our hypoth-
esis is that the theory of conservation of resources by psychologist S.E. Hobfoll
[13] offers an interesting lead. In this theory, the drive for the acquisition and
protection of resources is at the core of the dynamics which explains the stress or
well-being of an individual, and is even able to predict it. But at first it should
be understood that in this theory, the notion of resource refers to many types of
objects: social ones such as self esteem or caring for others, material ones such
as a car, or physiological ones such as energy. The main principle is that individ-
uals strive to protect their resources, and to acquire new ones. This model has
been developed originally for psychotherapeutic purposes, and resources con-
sidered are the ones which are critical in the life of an individual. But we believe
that it can be adapted for the context of realistic simulations. This framework
is generic to every computational environment where a description of available
resources is provided, along with the behaviours associated with the acquisition
or protection of these resources.

3 Proposed Model

3.1 Principle of the Model

The model is based on the following principles: (a) an agent strives to acquire
resources that it desires (b) once a resource is acquired, an agent tries to protect
it (c) an agent’s well-being depends on its capacity to acquire or protect resources
(d) an agent’s well-being regulates the tendency for acquisition or protection of
resources.

A desired resource triggers acquisition behaviours, and a threatened resource
triggers protective behaviours. Each resource type is associated with a partic-
ular set of acquisition and protective behaviours. For example an acquisition
behaviour for a resource “Position” in a waiting line could be “move forward”,
and an acquisition behaviour for a resource “Social Interaction” could be “talk to
somebody”. A protective behaviour for an acquired “Position” could be “move
forward just behind next agent”, and a protective behaviour for an ongoing “So-
cial Interaction” could be “speak loud” in order not to be interrupted by other agents.

A need value for a resource represents the level to which an agent wants the resource. This need level is dynamic and can change over time as a consequence of agents’ behaviour or environment events. For example a fire will trigger a high need level for security, and a long waiting time in front of a ticket counter before a train departure will increase the need for the resource “Train Ticket”.

3.2 Architecture

An agent’s affective architecture is composed of five affective sets which constitute the base of the model. The presence of a resource in one of these sets is a key factor influencing an agent’s behaviour. A distinction is made between a type of resource and an instance of resource. For example a type “TicketCounter” could have several instances in the simulation environment, e.g. ticket counters situated on the map with a location and effectively usable by an agent.

Let $A$ the set of agents and $R$ the set of resources in a simulation world $S$, with $A \subset R$. ∀$i$ ∈ $A$ at time $t$, we denote the four following affective sets:

- $N_i(t)$, the resource types that $i$ needs;
- $DR_i(t)$, the resource instances desired by $i$;
- $AR_i(t)$, the acquired resource instances of $i$;
- $TR_i(t)$, the threatened resource instances of $i$;
- $LR_i(t)$, the resource instances that $i$ has lost.

Let $V \in \mathbb{N}$ be a finite set of values. ∀$r$ ∈ $DR_i(t)$, we denote $\mu^r_i(t)$ ∈ $V$ the level of desire that $i$ has for a resource $r$ at time $t$. This value depends on the need value for the resource type of $r$ denoted as $\mu^\text{type}(r)_i(t)$, and the properties of $r$. For example if an agent $i$ needs a resource of type “Ticket Counter” in order to buy a resource of type “Train Ticket”, its desire for two instances $tc_1$ and $tc_2$ with $\text{type}(tc_1) = \text{type}(tc_2) = \text{TicketCounter}$ depends on: (i) $\mu^{\text{TicketCounter}}_i(t)$ (ii) the location properties of $tc_1$ and $tc_2$, since the closest is a ticket counter to $i$, the more it is interesting for $i$. Of course other factors are related to this situation, like the length of the waiting line to access a ticket counter instance, but we simplified this situation for explanation purpose.

Affective sets are initialized before the start of a simulation. It is possible to set them empty, to generate random desired resources, or to set them with specific resources in order to run a given scenario. Some needs for resource types like “Food”, “Drink”, “Social Image” or “Physical Integrity” should always be added for a realistic behaviour, unless no resource and no behaviour in the simulation environment allow to acquire or protect these resources. An example for a default setting could be: $\forall i \in A$, $N_i(t) = \{\text{Food, Drink}\}$, with $\mu^\text{Food}_i(0) = \mu^\text{Drink}_i(0) = x$, where $x$ corresponds to a need value. An example for a setting related to a scenario where an agent $i \in A$ has to buy a train ticket could be: $N_i(t) = N_i(t) \cup \{\text{TrainTicket}\}$, with $\mu^\text{TrainTicket}_i(0) = x$. 


$f(\text{TrainDepartureTime})$, where the more train departure time is close, the more the need for a train ticket is increased.

It is possible to set an agent’s personality in refining its need set. For example an agent could have a strong need for type as “Luxuous Car”, “Uncrowded Place”, “Social Interaction” or “Candy”.

∀ $r \in R$, there is compensation degree $C_i^r(t) \in [-V, V]$ which is the level to which a resource $r$ can decrease or increase $\mu_i^{\text{type}}(t)$. Two instances of type “Food” may not compensate agent’s need for food at the same level. Given two instances of type “Food” in the simulation environment which are hamburger and carrot, it is possible to set $C_i^{\text{hamburger}}(t) > C_i^{\text{carrot}}(t) > 0$. This means that the instance hamburger decreases $\mu_i^{\text{Food}}(t)$ with a higher degree than the instance carrot. It is possible to set individual characteristics for some agents in order that they react to the same instance in a different manner : some may satisfy their need for “Food” with the instance carrot whereas others may not.

During the simulation, addition and removal of resources in affective sets, as well as behaviour selection, are handled by the Affective Controller. This module takes into account resources perceived in agent’s environment, behaviours executed by other agents, and agent’s needs level. Each behaviour selected by this module has the purpose to acquire or protect a resource.

![Affective Controller Diagram](image)

**Fig. 1. General Architecture**

### 3.3 Behaviour Realization

The set of behaviours that can be performed by an agent $i$ includes the acquisition behaviours corresponding to the agent’s desired resources, and the protec-
tive behaviours corresponding to the agent’s threatened resources. Let $B_i(t)$ the set of behaviours that can be performed by an agent $i$ at time $t$. A behaviour $b \in B_i(t)$ has effects over resources during and after its realization for a given set of agents denoted as $\text{patients}(b)$. For example, if an agent $i$ performs the behaviour “insult” towards an agent $j$ during a verbal confrontation in a waiting line, the consequences of this behaviour is that $j$’s “Social Image” resource will be threatened, and this will trigger protective behaviours from $j$ in order to protect this resource.

\[ \forall b \in B_i(t), \forall j \in \text{patients}(b), \text{we denote :} \]

- $R^+_{ij}(j,t)$ : resource instances acquired by $j$ at time $t$;
- $R^-_{ij}(j,t)$ : resource instances of $j$ threatened at time $t$;
- $R^0_{ij}(j,t)$ : resource instances of $j$ protected at time $t$;
- $R^\rightarrow_{ij}(j,t)$ : resource instances lost by $j$ at time $t$;

These effects represent the agents’ knowledge upon the consequences of their behaviours. However these effects are not guaranteed, because each agent doesn’t know how other agents will react to a given behaviour. That means that if an agent $j$ engages in a protective action for its resource “Social Image”, this may lead to an aggressive physical reaction from the other agent, and this will have consequences for $j$ that are worse than the loss of its resource “Physical Integrity”.

To perform behaviour selection, a utility value is computed for each behaviour $b \in B_i(t)$, taking into account the behaviour’s effects described above. This value is computed with the compensation value of a resource upon an agent’s need level: a decrease of a need level is considered as a reward, and an increase is considered as a cost. Hence the loss of a resource like “Physical integrity” is a cost, since it causes an increase in agent’s need level for “Physical Integrity”: the agent no longer possesses the resource satisfying its need. The behaviour selected by an agent $i$ corresponds to the behaviour with the maximum positive utility for $i$.

See figure 1 for an overview of the general architecture of the model.

3.4 Individual Parameters

An agent knows its needs, the behaviours that it could realize in its environment on perceived resources, and the typical effects of these behaviours. It can therefore anticipate gains and costs. These raw values can be modified by individual factors such as agent’s well-being and egoism/altruism. The well-being of an agent acts as a sensor that guides an agent towards appropriate behaviours to readjust the state of its affective sets. For example an agent that has endured too many losses has a low well-being that pushes it to acquire new resources. That means that if an agent has lost an important resource such as its job, it may try to readjust its well-being with easy resource acquisitions like resource instances of “Food” type. Egoistic agents give more importance to their own payoff, and altruistic agents give more importance to other agents’ payoff.
4 Example

We consider a situation where agents have to buy train tickets provided by ticket counters in the simulation environment. Hence the provided resources are train tickets and ticket counters. Agents know that when a counter is occupied by another agent $a$, if they go to the waiting line it isn’t costly for $a$, whereas if they go directly to the counter it is costly for $a$. Indeed, when an agent is in front of a counter it considers that it has acquired the counter resource, which allows it to purchase a ticket resource. When another agent $b$ comes at the same time in front of a ticket counter, the counter acquired by $a$ is threatened. $a$ may choose to execute a protective behaviour in order to protect its resource, like telling $b$ to go away. When agents are in a waiting line, they each possess a position resource attached to this waiting line. The acquisition of positions in waiting lines in real life, as well as many other resources, are regulated by FIFO rule (First In, First Out)\cite{12}. When an agent $b$ ignores this rule so as to gain positions, it is costly for all agents between the current and previous position of $b$. If agent’s $b$ need for a train ticket is very strong and if $b$ is egoistic, it gives a great importance to the reward brought by a ticket acquisition, and a small importance to the positions lost for other agents. So $b$ can choose the behaviour of ignoring waiting line’s rule.

5 Discussion and Future Work

We presented in this paper an architecture aimed at providing virtual agents with affective behaviours in an urban simulation. Our hypothesis is that simulating behaviours that can be labelled as “emotional” do not necessarily requires an architecture grounded on emotional categories. Instead, we think that such behaviours are strongly motivated by resources. For example, even if behaviours in a waiting line are assumed to be driven by emotions, we tend to favor theories that considers a waiting line as a small social system regulated by the principles of property, that human beings are able to adopt naturally while not knowing the legal tenants of property \cite{12}. Our hypothesis is that the processes of resource acquisition and protection are the basis of affective behaviours. Actually, this idea applies in various contexts : in case we run from a fire, we try to protect our primary resource which is our body, in case we become friends with someone, it is because we find useful resources in this friendship (see the “social exchange theory” by Thibaut and Kelley \cite{15}), and so on. We believe that the architecture presented in this paper is generic, can account for several principles in social science and psychology theory, and is adaptable to many simulation contexts.

On another hand, the absence of emotional categories brings limitations. Our view is that emotions are concepts, and as concepts they are necessary for communication and reasoning. There are cultural patterns of facial expressions \cite{3} which are not currently mapped in the model of conservation of resources, and the emotional vocabulary is a large part human langage. A necessary work
has to be done on how the model of conservation of resources can be associated with concepts.

The model is currently under implementation for the context of a waiting line in front of a ticket counter. More contexts will have to be implemented in order to validate the genericity of the model. The evaluation will involve two procedures. First, agents’ behaviours in the simulation will be compared to behaviours described in psychological and sociological studies. Second, the credibility of agents’ behaviour will be rated by users after they have watched a simulation video. Finally once the evaluation protocol is completed, we plan to extend our model to groups and crowds, in order to use it for simulations with a large amount of agents.

References