Autonomy in Virtual Agents: Integrating Perception and Action on Functionally Grounded Representations

Argyris Arnellos[1], Spyros Vosinakis[1], George Anastasakis[2], and John Darzentas[1]

[1] Department of Product and Systems Design Engineering,
University of the Aegean, Hermoupolis, Syros, Greece
{arar,spyrosv,idarz}@aegean.gr
[2] Department of Informatics,
University of Piraeus, Greece
anastas@unipi.gr

Abstract. Autonomy is a fundamental property for an intelligent virtual agent. The problem in the design of an autonomous IVA is that the respective models approach the interactive, environmental and representational aspects of the agent as separate to each other, while the situation in biological agents is quite different. A theoretical framework indicating the fundamental properties and characteristics of an autonomous biological agent is briefly presented and the interactivist model of representations combined with the concept of a semiotic process are used as a way to provide a detailed architecture of an autonomous agent and its fundamental characteristics. A part of the architecture is implemented as a case study and the results are critically discussed showing that such architecture may provide grounded representational structures, while issues of scaling are more difficult to be tackled.

Keywords: Autonomy, representation, interaction, functional grounding, intelligent virtual agent, anticipation, belief formation.

1 Introduction

Intelligent Virtual Agents (IVAs) today play an increasingly important role in both fields of Artificial Intelligence (AI) and Virtual Reality (VR), for different reasons in each case [1]. The need for flexible and dynamic interactions between embodied intelligent entities and their environment as well as among themselves has risen quite early as the field of AI matured enough to set higher targets towards believable simulations of complex, life-like situations. Similarly, the added element of complexity, emerging as a natural consequence of realistic behaviour exhibited by embodied actors, either as units or in groups, and enhancing the feeling of user presence at an impressive rate, was soon recognized as the inevitable next goal of Virtual Reality. In effect, the two fields have converged towards a common aim as they concurrently progressed driven by their respective and, admittedly, quite different scientific concerns and individual goals. This has created new, fascinating possibilities, as well as a range of design and implementation issues [2].

Contemporary research in Virtual Environments has marked the need for autonomy in virtual agents. Autonomy has many interpretations in terms of the field it is being

J. Darzentas et al. (Eds.): SETN 2008, LNAI 5138, pp. 51–63, 2008.

used and analysed, but the majority of the researchers in IVEs are arguing in favour of a strong and life-like notion of autonomy, which should first of all replace omniscience in virtual worlds. As such, even from a practical perspective, autonomy is not a needless overhead. Since believability is considered as a crucial factor, virtual agents should appear to have limitations in their interaction with the environments, just as agents in the real world have. In this case, virtual agents should be able to interact with other agents and users in unexpected events and circumstances under fallible anticipations, to have limited perception capabilities and plausible action, to create and communicate new meanings about their environments and to exhibit novel interactions. Such agents could be used in dynamic and open-ended scenarios, where adaptability is needed.

As such, the notion of autonomy seems to play a very crucial role in the design of IVAs. Design and implementation benefits seem equally probable and significant. The IVA will be re-usable across a variety of different instances of a particular class of virtual worlds, not requiring re-engineering and additional implementation in case the virtual world model has to change. In general, behavioural flexibility in dynamic environments is an inherently-desired feature of any IVA design [3], [4]. Even more importantly, the process of generating the IVA's behaviour-control modules and modelling a particular IVA personality shall be disentangled at a great degree from the respective process of designing the virtual world and unconstrained by specific semantics – potentially unsuitable for cognitive processes – imposed by the virtual world's design, as predefined environment knowledge required would be reduced to a minimum [5]. However, for the time being, these issues remain theoretical to a large extent, reflecting the contemporary immaturity of the field and the diversity of the problems to be tackled. As it will be explained in the following sections, the main problem in the design of an autonomous IVA is that the respective models approach its interactive, environmental and representational issues as separate to each other, while the situation in biological agents, where one can find genuinely autonomous agents, is quite different.

In this paper an attempt towards the designing of autonomous IVAs is presented, at the theoretical, architectural and implementation level. Specifically, in section 2 a theoretical framework indicating the fundamental properties and characteristics of an autonomous biological agent is presented. Section 3 suggests the interactivist model of representations combined with the concept of a semiotic process as a way to model an autonomous agent and its fundamental characteristics and lays out a detailed architecture in order to implement such an autonomous IVA. A partial implementation of the proposed architecture is presented as a case study in Section 4. The conclusions of this work are mentioned in Section 5.

2 Designing Autonomous Artificial Agents

In almost any typical architecture of an artificial agent there are several components responsible for critical cognitive capacities, such as perceiving, reasoning, decision-making, learning, planning etc, regardless of whether the agent is situated in a virtual or in a physical environment (i.e. a robot). Interaction between IVAs and their environment is two-fold. IVAs sense their environment and generate knowledge about it

according to perception mechanisms, and they act upon their environment by applying actions generated by behaviour-control methodologies. Sensing and acting are achieved based on a processor of symbols, which in turn is connected to distinct modules of sensors and actuators/effectors. The processor relates together encoded symbols to form abstract representations of the environment. Each encoding results in a representational content which relates the cognitive system with the environment. Therefore, the respective environmental knowledge, as well as the generated actions is usually encoded in conceptual, symbolic forms, capable of expressing high-level, abstract semantics and compatible with the cognition processes supporting them. On the contrary, virtual environments are typically modelled according to low-level, platform-dependent symbols, which are quite efficient in both expressing detailed geometrical and appearance-related data, as well as effectively hiding deducible higher-level semantics [1], [6]. These approaches in modelling an IVA have been successful on tasks requiring high-level cognition, but they have been quite unsuccessful on everyday tasks that humans find extremely easy to manage. Additionally, such approaches cannot connect low-level cognition to higher-level behaviours, which is the key in the evolution of the autonomy of real-life biological agents. The problem is that in the respective architectures, the syntactic and semantic aspects of an IVA are separated, making the creation and enhancement of inherent meaning structures almost impossible. Particularly, reasoning and behaviour control involve relationships among generic, non-grounded representations of environment-related concepts, while sensory data retrieved and actions requested have to be explicit and grounded in order to be meaningful regarding a given instance of the underlying virtual environment. This is widely known as the symbol-grounding problem [7], which is a global problem in the design of artificial agents, and as such, it is also the most crucial problem of the merging between AI and VR inherent in IVAs, creating substantial implications at all levels of their design and implementation.

On top of that, the frame problem comes as a natural consequence. Specifically, since agent's functionality is based on predetermined and non-inherent representations, it will neither have the capacity to generalise its meanings in order to act on new contexts presenting similar relations and conditions, nor the ability to develop new representations and hence to function adaptively whenever is needed [8]. These problems have posed great difficulty in creating autonomous IVAs capable of successful interaction in complex and ill-defined environments. Agent's autonomy is compromised and belongs solely to its designer. In the realm of virtual agents, some attempts have been made to ground representations in the sensorimotor interaction with the environment [4], [9] but a cognitivist grounding theory should also explain the interdependence of each subsystem participating in the acquisition of the signal (i.e. the transducing system) with its environment and the central computational system, in order to be complete [10]. Everything that an IVA does, if it is to be an autonomous system, should, first of all, be intrinsically meaningful to itself.

In order to disentangle the designer from providing ad-hoc solutions to the interactive capabilities of an IVA, (through the introduction of pre-determined and ad-hoc semantics) one should try to see what the biological approach to autonomy –where symbol-grounding is not a problem and the frame problem is much more loose – may provide to the design of autonomous IVAs. The last two decades, there has been a growing interest in several theories of the development of autonomous biological

agents. A thorough analysis of these theoretical frameworks is out of the scope of this paper, but the reader may see [11] for a critical review, as well as for an analysis of an integrated framework of autonomous agents. What should be kept in mind is that autonomy and agency have many definitions with respect to the domain they are being used. Furthermore, autonomous systems are acting in the world for their self-maintenance, which is their primary goal.

For the purposes of this paper, agency is defined in a way that the suggested definition is more susceptible to an analysis of its functional characteristics. Therefore it is being suggested that a strong notion of agency calls for: *interactivity*, that is, the ability of an agent/cognitive system to perceive and act upon its environment by taking the initiative; *intentionality*, the ability of an agent to effect a goal-oriented interaction by attributing purposes, beliefs and desires to its actions; and *autonomy*, which can be characterized as the ability of an agent to function/operate intentionally and interactively based on its own resources. These three fundamental capacities/properties should be exhibited in a somewhat nested way regarding their existence and their evolutionary development and this makes them quite interdependent, especially when one attempts to understand if it is possible for each one of them to increase qualitatively while the others remain at the same level.

As such, it should be mentioned that there appears to be an interesting interdependence between the three fundamental properties, in the form of a circular connection between them. Specifically, Collier [12] suggests that there is no *function* without *autonomy*, no *intentionality* without *function* and no *meaning* without *intentionality*. The circle closes by considering meaning as a prerequisite for the maintenance of system's autonomy during its interaction. Indeed, this circle is functionally plausible due to the property of self-reference of an autonomous system. This is not just a conceptual interdependence. As it is analysed in [13], this is also a theoretical interdependence with a functional grounding, and as such, it sets some interesting constraints in the capacities that contribute to agency and it brings about some requirements in terms of the properties that an agent should exhibit independently of its agential level or in other words, of its level of autonomy. In short, these properties and their interdependence are considered as emergent in the functional organisation of the autonomous agent. The term 'functional' is used here to denote the processes of the network of components that contribute to the autonomy of the agent and particularly, to the maintenance of the autonomous system as a whole (see e.g. [14]). On the other hand, meaning should be linked with the functional structures of the agent. Hence, meaning should guide the constructive and interactive processes of the functional components of the autonomous agent in such a way that these processes maintain and enhance its autonomy. In this perspective, the enhancement of autonomy places certain goals by the autonomous system itself and hence, the intentionality of the system is functionally guiding its behaviour through meaning.

It should now be clear that the interactive, the environmental and the representational aspects of an autonomous agent cannot be separated to each other. This emergent nature of agency does not allow for the partitioning of agency in 'simpler problems' or/and the study of isolated cases of cognitive activity (e.g. perceiving, reasoning, planning, etc.). Nevertheless, these phenomena are quite typical in the research of autonomous artificial agents. However, the notion of a 'simpler problem'

should always be interpreted with respect to the theoretical framework upon which the design of the artificial agent relies.

As such, the primary aim of an attempt to design an autonomous virtual agent is not to design an agent that will mimic in a great detail the activities of a human. On the contrary, the aims of such research attempts should be the design of a complete artificial agent, that is, a design which will support, up to a certain satisfying level, the set of the abovementioned fundamental and characteristic properties of autonomy, by maintaining its systemic and emergent nature in different types of dynamically changing environments. The theoretical basis for such architecture, as well as the architecture itself is presented in the next section.

3 Virtual Agent Model

The design of an IVA following the principles sketched in Section 2 is based on the interactivist model of representation [15], [16], which favours more intentional and anticipatory aspects of the agent's representations. Due to space limitations a brief presentation of the model follows. In the interactivist model two properties are required for a system (and its functional subsystems) to be adaptable towards a dynamic environment. The system should have a way of differentiating instances of the environment and a switching mechanism in order to choose among the appropriate internal processes. In the system, several differentiating options should be available. These differentiations are implicitly and interactively defined by the final state that a subsystem would reach after the system's interaction with a certain instance of environment. One should be aware that although such differentiations create an epistemic contact with the environment, they do not carry any representational content, thus, they are not representations and they do not carry any meaning for the agent. What they do is that they indicate the interactive capability of system's internal process.

Such differentiations can occur in any interaction and the course of the interaction depends on the organization of the participating subsystem and of the environment. A differentiated indication constitutes emergent representation, the content of which consists of the conditions under which an interactive strategy will succeed in the differentiated instances of the environment. These conditions play the role of "dynamic presuppositions" and the respective representational content emerges in the anticipations of the system regarding its interactive capabilities. In other words, the interactive capabilities of the agent are constituted as anticipations and it is these anticipations that could be inappropriate and this is detectable by the system itself, since such anticipations are embedded in the context of a goal-directed system. It should be noticed that the possibility of internally detectable error on a functional basis, is the prerequisite for learning. Error guides learning in an autonomous system, where its capacity for directed interaction (towards certain goals) in a dynamic environment results in the anticipation of the necessity to acquire new representations [16].

In the proposed model, autonomy and the resulting intelligence is not considered as an extra module, but as an asset emerging from the agent's functionality for interaction. Specifically, the use of the proposed architecture aims at the unification of the modality of interaction, perception and action with the smallest possible number of representational primitives. What is missing is the way that the representations of the

anticipated interactions will be constructed in the system. In order to do this we use the form of the semiotic processes suggested by Peirce [17]. This semiotic framework is totally compatible with the theory of autonomous agents sketched above at the theoretical level [18]. At the design level of the suggested architecture the idea is that there is a *sign* (or *representamen)*, which is the form which a signal takes (and it not necessarily be material) and which designates an *object* (its *referent)* through its relation to the meaning (or *interpretant)* that it is created in the system while the latter interacts with the sign.

In this respect and for the purposes of the suggested architecture, the sign is everything that is comprised in the agent's perception of the environment (at any given moment of its interaction with it), the object are the set of the state of affairs present in the environment at the respective time (i.e. the agent's context) and the meaning is the functional connection between the sign and the object, which is being constructed by the agent during its interaction with the object. It should be mentioned that according to Peirce, meaning arise in the interpretation of the sign. Therefore, in our case, meaning depends on how and with what functional aspects (agent's internal processes) the form of the perceived sign is related to its object.

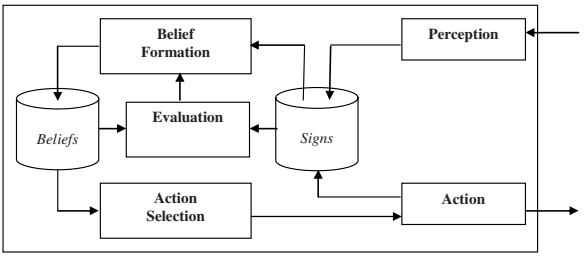

Fig. 1. The proposed agent architecture

Specifically, the resulting representations acquire a functional substrate, and as such, what really becomes represented in their contents is the outcome of the anticipated interactions internal to the agent. Since the agent interacts with the environment in order to achieve its goal and there is a way that the agent will understand whether its actions are towards a certain goal or not, then, the resulting representations necessarily, and by default, contain a model of the environment and of the way the agent may interact with it. Hence, the features of any object in the environment (material or not) are related to its functional properties for the agent. As a result, the agent perceives signs and constructs more developed signs as anticipations of successful interactions, with a functional substrate grounded on its organisation.

The agent architecture is presented in fig. 1. The design of the system is based on the assumption that the agent is equipped with a number of sensors and effectors that receive / send signals from / to the environment and with at least one internal status variable that describes the agent's degree of satisfaction. In the context of the proposed architecture, perception is defined as the process that reads values originating from the sensors and generates signs and action as the process of transmitting a value to one or more effectors at one time instance. Each time an action is performed, a respective sign is generated as well. The belief formation process transforms stored

signs into beliefs, i.e. the anticipated interactions of the agent, and the evaluation process continuously compares feedback from the actual sensors (incoming sings) to the anticipated interactions in order to detect errors and restructure these representations accordingly. Finally, the action selection process is the deliberative part of the agent that plans ahead and selects the action that is expected to bring the environment into a state that improves the agent's status. The individual processes of the proposed architecture will be explained in more detail in the next paragraphs.

3.1 Sign Production and Functional Relation

Signs are stored in the sign repository and contain all the information that the agent has collected during its interaction with the environment: the entities, properties, and events it has observed, the status changes it has noticed and the actions it has performed. The purpose of the sign collection is to process this data in order to generate beliefs about the environment and its dynamics. The detection of functional relations between signs and the formation of more complex signs, the generalization of similar signs into rules, and the formation and evaluation of hypotheses, will drive the meaningful interaction of an agent inside a dynamic environment.

We will present the notion of signs and the functionality of the architecture components that are utilizing them, through a simple example of a virtual environment. Let us think of an artificial environment, which contains agents and passive objects of various types. The environment is supported by a simple physically based modeling process: accelerated motion, collision detection and response, and friction. Agents can only perform the actions *move forward, rotate left* and *rotate right*, and therefore their only ability is to move around and to hit objects. Finally, a function defined by the designer updates the agent's status based on the state of the environment, e.g. the distance of a given object from a given region.

Let us define level-0 signs as any distinct property that the agent can observe. Each individual level-0 sign has a type (the property to which it refers) and a value. In our example, the level-0 signs could be:

- *time*: an integer value that increases by one after every interaction loop of the agent
- *object*: the unique id of an object
- *position*: an object's position in the 2D Cartesian space
- *orientation*: an object's orientation defined as the angle between its front vector and the x-axis.
- *action*: the type of action that the agent has performed
- *event*: any event of the environment that the agent can detect through its senses, e.g. a collision between two objects
- *status*: the internal status value of the agent.

Signs of level-1 are functionally grounded level-0 signs and are the ones generated by the agent's perception and action processes. In the above-mentioned example, four level-1 signs can be produced: The first is the *object-perception* sign that relates a time value t, an object o, a position (x,y) and an orientation r. The functional relation (semantics) of this sign is that object o at time t was located in (x,y) and had the orientation r. Object-perception signs are generated every time an agent observes an object

through its sensors, e.g. whenever the object is in its field of view. A special case of this sign is the one generated about the agent itself, which describes its own property values at every timeframe. The second sign is the *event-perception* sign that associates a time value *t* with an event *e*, stating that the event *e* was perceived at time *t*. The third sign is the *status* sign, which associates a status value *s* with a time value *t* stating that the value of the agent's internal status at time *t* was *s*. Finally, the fourth sign is the *action* sign, which associates an action *a* with a time value *t* to denote that the agent performed *a* at time *t*.

Multiple level-1 signs may contain the same level-0 sign. E.g. two object-perception signs that refer to the same object, or an event-perception and an object-perception that took place at the same time. In these cases, the level-1 signs can be functionally related based on the fact that they share one or more common properties, and form level-2 signs. Equivalently, two or more level-2 signs can be related if they contain one or more common level-1 or level-2 signs and form signs of level-3, etc.

3.2 Belief Formation, Evaluation and Action Selection

Apparently, from any given set of level-1 signs a huge number of functional relations leading to higher level signs can be produced. Some of these actually contain data that may be a useful basis for the agent's formation of meaning, provided that they are properly processed. The challenge for the agent's cognitive abilities is to be able to generate the appropriate signs, to detect similarities and to formulate hypotheses, in order to produce empirical laws that describe the environment's dynamics. However, the generation and comparison of all possible signs will face the problem of combinatorial explosion after very few interaction cycles. A possible solution to this problem is to have the designer insert a number of heuristic rules that will drive the production of higher level signs towards more meaningful representations. E.g. one heuristic rule could be to assign higher priority to the combination of two object-perception signs that refer to the same object and to successive time frames. The application of this rule will lead to higher level signs that describe *fluents*, i.e. object property values that change in time. E.g. such a rule could create level-2 signs that associate two successive positions of an object. This sign actually represents the velocity of the object at that time frame. Two such successive signs could be associated using the same heuristic rule and form a level-3 sign which will represent the acceleration of the same object.

One possible set of heuristic rules that can be employed in order to drive the belief formation process into representations that may assist the agent's deliberative behaviour is the following;

1. *fluents*: Two signs that include the same object sign and successive time signs are joined to form a fluent. The semantics of the fluent sign are that an objects property values v_1, v_2, …, v_k at the time point *t* have changed to v'_1, v'_2, …, v'_k at *t*+1.
2. *action effects*: A sign that includes the agent's property values at time t, an action sign at time t and a third sign with the agent's property values at t+1 can be joined to form a new sign that describes the effects of the agent's action at time t.

3. *event effects*: Similarly to actions effects, a sign that describes an event at time *t* can be connected to signs that include other objects' property values at times t and t+1.
4. *desired states*: A status sign at time *t* is connected to other signs that include the time point *t*. The aim of these signs is to lead to hypotheses about states of the environment that affect the agent's internal status value regarding its goals.

Even if a number of 'interesting' higher level signs can be generated using some heuristic rules, the information about past events and property changes of specific objects is of not much practical use to the agent. These signs will be useful if they can be generalized and lead to the formation of laws, which the agent will use to predict future states of the environment. In the proposed architecture, this sort of abduction is carried out by the belief formation process, which compares signs of the same type and attempts to generalize them. The process examines a series of signs of the same type and tries to detect patterns: which values remain unchanged, if the rate of change of a property value is constant or linear, etc. The generalization attempts to replace the level-0 signs into variables and assumes that the same belief holds for all possible values of each variable, e.g. for all time points *t*, for all objects *o*, etc. Additionally, in the cases of fluents, action effects and event effects, it is assumed that for each variable v'_i that changed value at t+1 there is a function f_i for which holds: $v'_i - v_i = f_i(v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k)$, where $v_1, v_2, \ldots, v_k$ are the variable values at time *t*. If the belief holds and this function can be properly approximated based on previous observations, the agent will be able to predict the next state of (part of) the environment based on its current state. A simple approximation of a function for a given set of input variables is to detect the nearest neighbors of the input variable set and to perform an interpolation on the output variable based on the Euclidean distance from the neighbors.

Let us present a simple example of generating beliefs. A generated action-effect sign could associate the agent's own property values (object-perception sign) at time *t*, the action it performed at time *t*, and the agent's property values at time t+1. Let *p, r* be the agent's position and rotation angle at time *t* respectively, *a* be the action it performed, and *p', r'* be the agent's position and rotation angle at time t+1. If a belief is formed that the agent's position and angle are affected by each one of its actions, then the belief formation algorithm will have to approximate the functions f_1 and f_2, where $p' - p = f_1(r, a)$ and $r' - r = f_2(p, a)$. The constant inspection of action-effect signs will provide more samples of f_1 and f_2 and will lead to the better approximation of the functions. Assuming that the agent is rotating at a constant rate and that it is moving forward by a constant distance, the comparison of a multitude of samples will result to the following values: if *a* is an action *rotate left* or *rotate right*, f_1 will be always equal to zero and f_2 will be equal to a constant number, whilst if *a* is a *move forward* action, f_1 is not constant and will have to be approximated and f_2 will be equal to zero.

The beliefs generated by the belief formation process are subject to failure, as they are the product of abduction and may not be true in all contexts. E.g. a belief about the implications of a collision on the motion of two objects may be true for objects of similar mass, but will fail if one of the objects is significantly heavier, or even inanimate. The aim of the belief evaluation process is to test the validity of the agent's beliefs by examining the existing and the incoming signs and to determine whether they support the beliefs or not. If a sign is detected that does not support a belief, the

sign is treated as an exception, and the belief is being restructured in an attempt to exclude the exception. In that case, one of the variables of the belief is selected, and the respective value of the exceptional sign is excluded from it. However, since this variable is randomly selected, the result of the restructuring may still be subject to failures. Belief generation and evaluation is a continuous iterative process that drives the agent's representations towards more meaningful functions.

The distinction between cases that a belief applies and others that it does not, leads to the introduction and formation of categories concerning the values of perceivable properties. E.g. an action-effect belief that is until now proven to be applicable only to a subset of perceivable objects, categorizes objects. If a category proves to be persistent in time, it may describe the value of an essential property of the environment, which is perceived indirectly by the agent, e.g. mass, friction, velocity, etc. In such a case, the agent generates a symbol that describes the category. The formation of new symbols and signs will ultimately to lead to a higher layer of abstraction.

The action selection process decides about the next agent action to perform on the basis of maximizing the agent's status value, i.e. the agents 'satisfaction' as defined by the designer. Assuming that the agent has formulated a number of beliefs concerning fluents and events of the environment, states that increase its status value, and effects of its actions on its own property values, the agent can generate plans with a simple forward chaining algorithm in order to select the action that is expected to lead to the achievement of the its goal. The algorithm will estimate the effects of applying every possible action in the current time frame using the functions obtained in the respective beliefs and will estimate all the possible states of the environment in the next time frame. Using this process repeatedly up to a maximum number of states, a plan of action can be generated. It is, however, possible that the plan will fail because the beliefs that determine the effects of actions and the evolution of the environment are based solely on the agent's observations and approximations and will never grasp the actual 'physical' laws of the environment in their entirety. If the agent does not possess enough beliefs to construct plans, it may select actions that are expected to increase its knowledge.

4 Case Study

A partial implementation of the proposed architecture in the context of a simple virtual environment with agents, passive objects, and a collision detection and response mechanism is presented as a case study (fig. 2.a). Initially, the agents are unaware of the function of the actions they can perform. Consequently, they observe the effects of their own actions in an indirect fashion. In addition, agents are capable of building action-effect signs as their sign repository grows over time, and to transform them into beliefs. Their internal status value (satisfaction) increases as they minimize their distance from a specific target in space. In the experiments performed, targets could be set to be static points or actual objects in the environment. The latter lead to constant re-locating of targets as the agent hits objects when it collides with them. The agent is equipped with a field-of-view sensor, thanks to which it can receive signals describing locations of objects in the world that it stores as signs in its repository. During its interaction loop the agent tries to select the most effective action towards its target based on the beliefs it has generated concerning the effects of its actions.

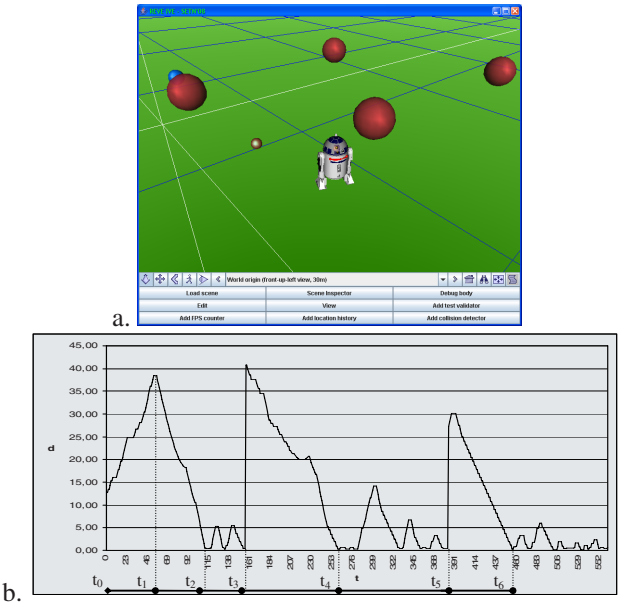

Fig. 2. a. Virtual world application **b.** agent's distance to target over time

Experiments were focused on an analysis of the agent's behavior in terms of deliberate modification of its own location over time by means of the available actions (move forward, rotate left and rotate right), with respect to target locations. The diagram in fig.2b above illustrates a sampling of 561 readings of the agent's Euclidean distance to a specific target on the x-z plane over time. As shown on the diagram, the agent's initial lack of anticipated interactions drive it to move in a random fashion, resulting in the increase of its distance to the target over time. While doing so, however, it progressively refines its representations as it observes the effects of its random actions. As a result, approximately at $t_1 = 56$, the representations are accurate enough to enable the agent to actually plan its motion towards the target. Around $t_2 = 112$, the agent has reached very close to the target, and continues to refine its representations by moving around it, being able, however, to restrain its motion in a small region. At $t_3 = 155$ the agent is assigned a new target, hence the steep increase in the distance readings. In the samples to follow, however, it is clear that its motion towards the target is now more controlled and directed than before, as the distance constantly drops, even though not at a constant rate. At approximately $t_4 = 259$ the agent has reached the area of the target and behaves in a similar fashion as in (t_2, t_3), maintaining a reasonably small distance to the target. At $t_5 = 381$, the agent is assigned yet another target; this time, its motion until it reaches it at $t_6 = 453$ is almost linear, and the distance drops at a constant rate. In conjunction with the agent restricting its motion to an even smaller region of the target after t_6, the above results indicate the agent's capability to progressively refine its beliefs about the effects of its own actions and, eventually, put them to actual, deliberate and effective use as a means to achieve its goal.

5 Conclusions

We have presented an example as an application of the proposed framework, where agents create grounded representational structures based on their interaction with the environment. In the proposed model, autonomy and the resulting intelligence is not considered as an extra module, but as an asset emerging from the agent's functionality for interaction. It has to be noted that relevant attempts have been made also in [19], [20] and [21] with a much more specific world for the agent to interact with and with many more hand-coded aspects of its functionality. Specifically, what has been achieved in this paper is that the logic of semiotic sign processes provides a tool for the integration of perception and action under functional representations, which refer to certain environmental states of affairs as subjective anticipations of possible outcomes of the interaction of the agent with the respective instance of the environment. While the way differentiations of the environment and the respective indications for possible outcome of an interaction (i.e. anticipations/beliefs) are formed in a way compatible with the one suggested in the proposed theoretical framework – all under the guidance of a certain goal which is implicitly defined through a graded satisfaction signal, as in any biological agent – the number of introduced abstraction levels (sign levels – more developed signs and anticipations which result in more complex generalisations) and the criterion for their introduction are inevitably hard-coded. Nevertheless, keeping in mind that the aim of this paper was to set the basis for the designing of an IVA which will support, up to a certain satisfying level, the fundamental properties of autonomy, in different types of dynamically changing environments, we have demonstrated that the suggested architecture is on the right track, while the abovementioned difficulties remain as a significant challenge in any relevant attempt.

References

1. Aylett, R., Luck, M.: Applying Artificial Intelligence to Virtual Reality: Intelligent Virtual Environments. Applied Artificial Intelligence 14(1), 3–32 (1999)
2. Thalmann, D.: Control and Autonomy for Intelligent Virtual Agent Behaviour. In: Vouros, G.A., Panayiotopoulos, T. (eds.) SETN 2004. LNCS (LNAI), vol. 3025, pp. 515–524. Springer, Heidelberg (2004)
3. Gillies, M., Ballin, D.: Integrating Autonomous Behavior and User Control for Believable Agents. In: 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS 2004), pp. 336–343 (2005)
4. Dinerstein, J., Egbert, P.K.: Fast multi-level adaptation for interactive autonomous characters. ACM Transactions on Graphics 24(2), 262–288 (2005)
5. Gratch, J., Rickel, J., Andre, E., Badler, N., Cassell, J., Petajan, E.: Creating interactive virtual humans: Some assembly required. IEEE Intelligent Systems 17(4), 54–63 (2002)
6. Kasap, Z., Magnenat-Thalmann, N.: Intelligent Virtual Humans with Autonomy and Personality: State-of-the-Art. Intelligent Decision Technologies 1(1-2), 3–15 (2007)
7. Harnad, S.: The Symbol Grounding Problem. Physica D 42, 335–346 (1990)
8. Janlert, L.E.: Modeling change: The frame problem. In: Pylyshynn, Z.W. (ed.) The robots dilemma: The frame problem in artificial intelligence. Ablex, Norwood (1987)

9. Rickel, J., Johnson, W.L.: Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. Applied Artificial Intelligence 13, 343–382 (1999)
10. Ziemke, T.: Rethinking Grounding. In: Riegler, P., von Stein (eds.) Understanding Representation in the Cognitive Sciences. Plenum Press, New York (1999)
11. Arnellos, A., Spyrou, T., Darzentas, J.: Towards the Naturalization of Agency based on an Interactivist Account of Autonomy. New Ideas in Psychology (Forthcoming, 2008)
12. Collier, J.: Autonomy in Anticipatory Systems: Significance for Functionality, Intentionality and Meaning. In: Dubois, D.M. (ed.) The 2nd Int. Conf. on Computing Anticipatory Systems. Springer, New York (1999)
13. Arnellos, A. Spyrou, T., Darzentas, J.: Emergence and Downward Causation in Contemporary Artificial Agents: Implications for their Autonomy and Some Design Guidelines. Cybernetics and Human Knowing (Forthcoming, 2008)
14. Ruiz-Mirazo, K., Moreno, A.: Basic Autonomy as a Fundamental Step in the Synthesis of Life. Artificial Life 10, 235–259 (2004)
15. Bickhard, M.H.: Representational Content in Humans and Machines. Journal of Experimental and Theoretical Artificial Intelligence 5, 285–333 (1993)
16. Bickhard, M.H.: Autonomy, Function, and Representation. Communication and Cognition — Artificial Intelligence 17(3-4), 111–131 (2000)
17. Peirce, C.S.: The Essential Peirce. Selected Philosophical Writings, vol. 1(1867–1893). Indiana University Press, Bloomington, Indianapolis (1998)
18. Arnellos, A., Spyrou, T., Darzentas, J.: Dynamic Interactions in Artificial Environments: Causal and Non-Causal Aspects for the Emergence of Meaning. Systemics, Cybernetics and Informatics 3, 82–89 (2006)
19. Tani, J., Nolfi, S.: Learning to perceive the world as articulated: An approach for hierarchical learning in sensory-motor systems. In: Proceedings of the fifth international conference on simulation of adaptive behavior. MIT Press, Cambridge (1998)
20. Vogt, P.: The emergence of compositional structures in perceptually grounded language games. Artificial Intelligence 167(1-2), 206–242 (2005)
21. Roy, D.: Semiotic Schemas: A Framework for Grounding Language in the Action and Perception. Artificial Intelligence 167(1-2), 170–205 (2005)