The Art and Science of Synthetic Character Design

Christopher Kline and Bruce Blumberg
Synthetic Characters Group, MIT Media Laboratory
20 Ames Street, Cambridge, MA 02139
{ckline,bruce}@media.mit.edu

Abstract

Drawing from ideas in both traditional animation and modern philosophy, we present a methodology for designing synthetic characters. The goal of our approach is to construct *intentional* characters that are both compelling, in the sense that people can empathize with them, and understandable, in that their actions can be seen as attempts to satisfy their desires given their beliefs. We also present a simple, value-based framework that has the flexibility to implement the subsystems necessary for the construction of intentional characters.

1 Introduction

One of the most promising intersections of entertainment and AI research is in the creation of believable *synthetic characters*—simple but complete three-dimensional situated agents who can do and express the right things in a particular situation or scenario. Examples of these types of agents include non-player characters in computer games, digital 'extras' in Hollywood movies, and computer-based artificial pets. Often these characters do not need to perform complex reasoning about the world or build intricate plans to achieve difficult goals. Instead they may effectively play out their roles by reacting to internal and external influences in ways that are both predictable and consistent with the scenario for which they were designed.

In the process of learning to build these types of characters we have often found ourselves struggling with two fundamental problems. First, what kinds of properties or qualities have we, as observers, come to expect from a believable character? Second, given these expectations, what is the 'right way' to go about implementing them?

This paper presents an overview of the lessons we have learned from our experience in building several complex synthetic characters. We begin by discussing a theory of how people go about understanding characters and then identify some subsystems that we have found to be important in building characters that are compelling and easy to understand. Next, we overview several approaches to these subsystems and show how, by separating out the semantic differences of these approaches, we can arrive at the basic activity of each. We then describe a simple value-based framework we have developed for character construction, showing how each subsystem can be implemented with the four components of our framework. Finally, we conclude with some results from our experiments with this framework and suggest directions for future exploration.

2 Expectations of a Synthetic Character

To learn how to build believable characters we look back upon the rich history of traditional character animation. When looking at a character brought to life by a great animator we know exactly what that character is thinking and feeling at every instant and, while we may not know exactly what it is about to do, we can always call upon our perception of its desires and beliefs to hazard a guess. Even when our guess is wrong, the resulting behavior nearly always "makes sense".

Classics like *The Illusion of Life* (Thomas 1981) explain the art of creating believable characters, which is fundamentally the art of revealing a character's inner thoughts—its beliefs and desires—through motion, sound, form, color and staging. But why do these techniques work? The American philosopher Daniel Dennett believes that they work because, in order to understand and predict the behavior of the animate objects around them, people apply what he calls the *intentional stance* (1987). The intentional stance, he argues, involves treating these objects as "'rational agents" whose actions are those they deem most likely to further their 'desires' given their 'beliefs'" (1998).

Desires are the key to understanding and identifying with a character. When we see the wolf look "longingly" at Little Red Riding Hood, perhaps licking his lips, we conclude that the wolf is hungry and wants to eat our heroine. How do we arrive at this conclusion? By applying the intentional stance, of course! Why else would he be acting hungry unless he *was* hungry?

Beliefs are what turn desires into actions, reflecting influences such as perceptual input ("If I see a stream, then I believe I will find water there"), emotional input ("Because I am afraid of that person, I will run away from him"), and learning ("The last time I was in this field I saw a snake, therefore I will avoid the field today"). We understand the actions of characters by inferring how their beliefs influence the ways they attempt to satisfy their desires.

How can we apply both the insights of skilled animators and knowledge of the intentional stance to build a synthetic character that people find compelling *and* understandable? From the standpoint of engineering, we can break these expectations down into a short list of functional subsystems:

- Motivational drives
- Emotions
- Perception
- Action selection

2.1 Motivational Drives

For a character to appear properly motivated it must continue to work towards satisfying its desires while gracefully handling unexpected situations. For example, a creature that is starving may temporarily ignore its hunger in order to flee from an approaching predator. Once the danger has passed, however, the creature should resume searching for food. By biasing action selection towards behaviors that will satisfy the internal needs of the creature, motivational drives provide a way to achieve goal-oriented behavior.

Several researchers have addressed the problem of motivations in the context of building creatures. One example is the work of Blumberg (1996), who used temporally cyclic 'internal variables' in the design of a virtual dog to bias action selection and facilitate external direction of synthetic actors. In another domain, Breazeal (1998) has developed a motivational system for regulating interactions between a robot 'infant' and its human caretaker, with the goal maintaining an environment suitable for learning.

Most approaches agree on the general behavior of drives. Most importantly, they are cyclical and homeostatic—positive or negative deviations over time from the base state of 'satisfaction' represent under- and over-attention, respectively, to a corresponding desire. These desires can be attended to by the successful execution of attentive behaviors like eating, or by changes in external stimuli, such as temperature fluctuations or interactions with other creatures. When unattended to, drives slowly increase over time; the effect of attentive actions is to shift the value of the drive back towards its homeostatic base state.

2.2 Emotions

Emotions bias action selection in much the same way as drives. For example, a creature that is angry may be more prone to violent behavior than one who is happy. However, emotions also bias the quality of the character's motion. If the creature is sad it should walk sadly; if it is fearful it should reach for objects in a manner which conveys its fear. In this way emotion helps observers to form an empathic bond with the creature and makes its behavior appear properly motivated (Thomas 1981).

There are many approaches in the literature to the modeling of emotions and other affective phenomena. In so-called 'appraisal' theories of emotion the individual is said to make a cognitive appraisal of their current state relative to a desired state. For example, Reilly (1996) proposes that fear might be modeled as proportional to "the likelihood of failing to achieve the current goal" multiplied by "the importance of not failing". Others such as LeDoux (1996) argue that emotions can act at a level far below the cognitive, since animals can feel emotions without consciously understanding why. Combining these approaches, Velasquez (1998) presents a framework that models how emotional systems interact with the perceptual, motivational, behavioral, and motor systems.

The general consensus of these models is that, instead of increasing slowly over time as do drives, emotions typically exhibit a large impulse response followed by a gradual decay back down to a base state. By altering the decay term and the gains on stimuli one can adjust the magnitude and slope of the impulse response, shaping the characteristic response of the emotion. Adjusting these parameters across the space of emotions is equivalent to shaping the 'temperament' of the creature. Similarly, by altering the bias term on each emotion predisposes the creature to a particular emotional state, setting its 'mood'. These decay, bias, and stimulus terms represent the influences of a variety of systems[1], which in turn are affected by the current emotional state.

It is perfectly appropriate to model the influences of multiple emotions upon internal processes such as action selection, but it is difficult for human observers to visually perceive more than one emotion at a time. This is why animators tend to emphasize the most important emotion of a character, avoiding "mixed emotions". Because we are designing characters for humans to interact with, it is important for the underlying emotional model to support some notion of a 'dominant' emotion. This dominant emotion can then be used to parameterize motion and expression, giving the observer insight into the internal desires and beliefs of the character.

One example of such a parameterization is the animation system of Rose, Cohen, and Bodenheimer (1997), in which motor commands are specified in terms of verbs ("walk", "reach-for") and adverbs ("sadly", "impatiently"). Through the use of multi-dimensional interpolation, this system can be used to continuously modify a character's motion in order to represent the changing state of one or more emotions (for example, making a character move as if it is mostly happy, but slightly impatient and somewhat tired).

2.3 Perception

Fundamentally, a situated, embodied agent needs a way to "make sense" of the world in which it is situated. By this we mean two things. First, the creature needs a method of sensing the world around it; second, it must have a mechanism for evaluating the salience of incoming sensory information. The combination of a sen-

[1] E.g., factors include the neurobiological (e.g., hormones), motivational (intense hunger), cognitive (an impending conference deadline; the perception of a predator), and sensorimotor (posture)

sory stimulus and its corresponding evaluation mechanism is known as a *perceptual elicitor* or what ethologists (Lorenz 1973, McFarland 1993) refer to as a *releasing mechanism*.

Sensory information can be provided to a synthetic creature many forms, most of which fall into the three basic categories: real-world physical sensing, synthetic vision, and direct sensing. Physical devices like the temperature sensors in the motors of the Cog robot (Brooks 1996) and the infrared sensors on the mobile robots of Mataric (1994) are typical of real-world sensors. Synthetic vision techniques attempt to extract salient features from a physical scene rendered from the viewpoint of the creature; examples include the ALIVE system of Maes (1996) and the artificial fish of Tu and Terzopolous (1994). In direct sensing, creatures gain information by directly interrogating the world or an object within the world include; this is the approach taken by the boids of Reynolds (1987) and many video games.

One of the important contributions of Blumberg (1996), building on ideas from Lorenz (1973), Baerends (1976), and McFarland (1993), is the notion that external perceptual influences must be reduced to a form that is compatible with internal influences such as motivations and emotions. Using a consistent internal "common currency" is essential for addressing the issue of behavioral relevance—a piece of rotting food should be as compelling to a starving creature as a delicious-looking slice of cake is to a creature that has already eaten too much. Given this representational paradigm, opportunistic behavior is simply a side effect of the relative difference in weighting between external and internal influences.

2.4 Action Selection

Regardless of the particular implementation, the fundamental issues for any action selection scheme to address are those of adequacy, relevance, and coherence (Brooks 1990). Adequacy ensures that the behavior selection mechanism allows the creature to achieve its goals. Relevance, as noted above, involves giving equal consideration to both the creature's internal motivations and its eternal sensory stimuli, in order to achieve the correct balance between goal-driven and opportunistic behavior. Coherency of action means that behaviors exhibit the right amount of persistence and do not interfere with each other or alternate rapidly without making progress towards the intended goal (i.e., behavioral aliasing).

In an effort to achieve these goals in noisy and dynamic environments, the last two decades of agent research have seen a shift away from cognitivist 'Planning' approaches towards models in which behavior is characterized by the dynamics of the agent-environment interaction. In these environments, *nouvelle AI* researchers argue, collections of simple, competing behaviors that are tightly coupled with sensors and actuators can be more effective than complex planning mechanisms, while exhibiting many of the same capabilities. Examples of these approaches include the Pengi system of Agre and Chapman (1987), the subsumption architecture of Brooks (1986), the spreading activation networks of Maes (1991), and the "Society of Mind" theories of Minsky (1988).

In an attempt to leverage the advantages of both approaches, some hybrid systems like that of Firby (1987) have used a planner to make high-level behavioral decisions while using a reactive system for low-level control during behavior execution.

Inspired by ethological theories of behavior, some systems use a hierarchical organization to break complicated tasks down into specialized cross-exclusion groups (Minsky 1988) in which mutually-exclusive behaviors compete for dominance, using mutual and lateral inhibition to control arbitration (Ludlow 1976). These include most notably the Hamsterdam system of Blumberg (1994) and the work of Tyrrell (1993).

3 A Value-based Framework

In the previous section we talked about some of the important building blocks of a character that acts and emotes in a way that people find understandable and compelling. But how should one go about implementing these subsystems? In our experience we have found it useful to try a variety of approaches; this continual improvisation is made easier when the underlying framework makes it easy to implement and integrate different models.

The traditional approach to building creatures has been to focus on each of these subsystems individually. However, if we step back for a moment and consider them as a whole, two important regularities become apparent. First, there is a high degree of interdependence among subsystems—perception, emotions, and drives influence action selection, and the results of action selection in turn affect the external state of the world and the internal state of the creature. Second, the function of each can be interpreted as a quantitative mechanism. For example, the changing value of emotions and drives indicate the state of internal needs, perceptual elicitors determine the relevance of percepts, and action selection mechanisms choose the most appropriate behavior from among multiple competing ones.

What this suggests is that there is a great deal of common functionality among these subsystems. In many cases the functions performed by these subsystems can be seen as *simply different semantics applied to the same small set of underlying processes*. Consequently, instead of struggling to integrate multiple disparate models for each subsystem, it makes more sense to build them all on top of a framework that provides these shared constructs.

3.1 The Four Components

We have constructed this type of framework from four basic underlying components. The coherency of our framework comes from the fact that our primary internal representation is the floating-point value. In addition to being an intuitive way to think about emotions, drives, and sensory input, value-based frameworks have a number of other advantages. They are relatively easy to implement and fast at run-time, have useful parallels with reinforcement learning and neural networks, and are easily extendable because external semantics are kept separate from internal representation.

Granted, not everything is best represented numerically. However, for the purposes of getting along in the world, the processes which could potentially produce non-numeric representations (sensing and cognition, e.g.) can be seem as means to one end—action. And before any creature takes action it must first decide what action to take, which is a qualitative evaluation. Therefore, for the purposes of action selection, all semantic representations in our system are first converted to a value.

3.1.1 Sensors

In our system, the sensor primitive is an abstract component that operates on arbitrary input and outputs a set of objects appropriate to the sensor's functional criteria. Sensors typically use the external world or the internal state of the character as input. In addition, they may use the output of a different sensor as input; in this manner a directed, acyclic data-flow sensing network may be formed. For example, a `VisibleObjectSensor` could find all the visible objects in the world (through direct sensing, computational vision, or any arbitrary method), passing its output to a `DogSensor` to filter out everything but dogs.

3.1.2 Transducers

The transducer primitive operates on a set of input objects to produce a single floating-point output; transducers are the gateway through which sensor data enters the computational substrate. The values produced by transducers are often objective and the result of basic computations, such as the distance to the first sensed object. However, there is nothing to restrict a transducer from returning a subjective result from a complex computation—reasoning with predicate calculus about a set of input obstacles and returning the best heading in which to move, for example. Chains of sensors and transducers form the perceptual elicitors that allow the creature to react to internal and external situations.

3.1.3 Accumulators

The third primitive in our framework, the accumulator, is the primary unit of computation. Its inputs and gains are typically the output of transducers or other accumulators, and by constructing feedback loops it is possible to create highly connected networks which exhibit useful temporal behavior. The value V_t of an accumulator at time t for N inputs and gains is:

$$V_t = \sum_{i=0}^{N-1} \text{input}_{t,i} \cdot \text{gain}_{t,i} \quad (1)$$

where N is arbitrary.

3.1.4 Groups

The fourth primitive, the group, is used to organize accumulators into semantic groups and impose arbitrary behavior upon them. For example, a group might force the value of its accumulators to be zero except for the accumulator with the highest value. This abstraction keeps the syntax and configuration of the accumulators independent of their group semantics.

3.2 From Components to Subsystems

As an illustration we will now show one way in which each subsystem can be constructed from the components of our framework.

3.2.1 Drives

Motivational drives can be expressed using an accumulator with a feedback loop whose gain is at least one. Attentive and aggravatory stimulus inputs are given negative and positive gains, respectively, and one additional input-gain pair represents the magnitude of the growth term. The setup in Figure 1 creates a drive in the style of Breazeal (1998).

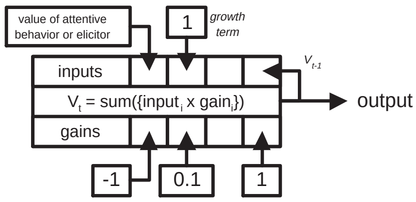

Figure 1: An accumulator-based motivational drive

Assuming that each stimulus$_i$ is a positive-valued stimulus working to satiate the drive, this configuration increases in value over time from a homeostatic base state of zero, according to (2).

$$V_t = V_{t-1} + \text{growth}_t - \sum_i \text{stimulus}_{t,i} \quad (2)$$

3.2.2 Emotions

Emotions can be implemented with a configuration similar to that used for drives where, instead of acting as a growth term, the input-gain pair biases the homeostatic base state. By limiting the gain on the feedback loop to the range $(0, 1)$ we can effect a gradual decay over time in the value of the emotion. This configuration, show in Figure 2, varies in time according to (3).

Often it is useful to organize emotions into cross-exclusion groups for the purposes of identifying the dominant emotion. By adjusting the inhibition between the competing emotions we can tailor the personality of the creature—making a fearful creature less prone to happiness, for example.

$$V_t = (V_{t-1} \cdot \text{decay}_t) + \text{bias}_t + \sum_i \text{stimulus}_{t,i} \quad (3)$$

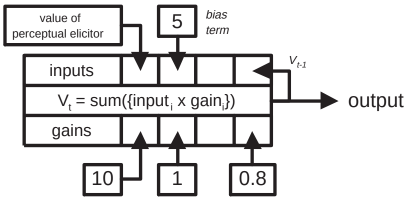

Figure 2: An accumulator configured as an emotion

3.2.3 Action Selection

A behavior is simply an accumulator that is semantically associated with a particular behavioral routine that it executes while 'active'; typically this involves sending a message (e.g., "walk") to an underlying motor system. Their inputs are the outputs of emotions, drives, and perceptual elicitors; whether a behavior is considered 'active' or not is determined by the semantics of its associated group. For example, autonomic behaviors like breathing and blinking might be contained in a group whose policy is to activate any behavior with a value above a certain threshold.

To achieve ethologically-inspired action selection policies, mutually exclusive behaviors can be organized into groups with cross-exclusion and mutual inhibition semantics and forced to 'compete' on the basis of their output values. Hierarchical action selection in the style of Blumberg (1996) and Tyrrell (1993) is easily implementable by associating each behavior with a reference to another group.

This method of implementing action selection has the advantage of making behavior design independent of action selection policy, allowing the designer to use the same behavior in many different contexts. For example, under normal circumstances a character might execute a swallowing behavior at regular intervals; this same behavior, however, might be a sub-behavior with an explicit order in the context of eating a meal. In our framework the same behavior can be used in both situations without requiring the designer to implement or have *a priori* knowledge of policy-specific details (e.g., connections to parent behaviors, execution order, etc.). This flexibility facilitates creating libraries of generic behavioral routines from which a variety of characters can be constructed.

4 Future Work

There are a many areas in which our system could be extended or improved. Most pressing is the need for better character design tools. The framework we have presented was intended as a kind of assembly-level language for building the various components of a complete character. While this flexibility has proven useful and valuable, it is currently tedious to construct complex characters in this fashion. We are currently looking into the development of a high-level behavior language or graphical interface from which we could compile the low-level internal representations discussed in Section 3.

Another area that we intend to pursue is the incorporation of learning. Though we have not yet implemented this in our existing characters, given the similarities between our work and the Hamsterdam system of Blumberg (1996) we are confident that our framework will accommodate a similar model of adaptation.

Figure 3: Two of the characters in *Swamped!*

5 Conclusion

Drawing from ideas in both traditional animation and modern philosophy, we have presented a methodology for designing synthetic characters. The goal of our approach is to construct *intentional* characters that are both compelling, in the sense that people can empathize with them, and understandable, in that their actions can be seen as attempts to satisfy their desires given their beliefs. We also presented a simple, value-based framework that has the flexibility to implement the subsystems necessary for the construction of intentional characters.

The concepts presented here were used to successfully build the many autonomous and semi-autonomous characters in *Swamped!*, an interactive cartoon experience premiered at SIGGRAPH 98. In this exhibit the participants use a *sympathetic interface* (Johnson 1998) to influence the behavior of a chicken character, with the intent of protecting the chicken's eggs from being eaten by a hungry raccoon. The raccoon character is arguably one of the most complex fully autonomous synthetic character built to date, comprised of 84 distinct behaviors influenced by 5 separate motivational drives and 6 major emotions. In addition, the continuously changing emotional state of the raccoon is conveyed through dynamically interpolated character motion and facial expressions.

Acknowledgements

The authors would like to thank Michal Hlavac, Michael Patrick Johnson, Kenneth Russell, Bill Tomlinson, and Song-Yee Yoon of the Synthetic Characters Group, and Andrew Wilson of the Vision and Modeling Group. Their experience, criticism, and hard work were instrumental in the design and implementation of our architecture.

References

P. Agre, D. Chapman. Pengi: An Implementation of a Theory of Activity. In *Proceedings of AAAI-87*. 1987.

G. Baerends. On Drive, Conflict and Instinct, and the Functional Organization of Behavior. In *Perspectives in Brain Research 45*. 1976.

B. Blumberg. Action-selection in Hamsterdam: Lessons from Ethology. In *From Animals to Animats: Proceedings of the Third International Conference on the Simulation of Adaptative Behavior*. 1994.

B. Blumberg. Old Tricks, New Dogs: Ethology and Interactive Creatures. Ph.D. thesis, MIT Media Lab. 1996.

C. Breazeal (Ferrell). A Motivational System for Regulating Human-Robot Interaction. In *Proceedings of AAAI-98*. 1998.

R. Brooks. Prospects for Human Level Intelligence for Humanoid Robots. In *Proceedings of the First International Symposium on Humanoid Robots (HURO-96)*. 1996.

R. Brooks. A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation* RA-2: 14–23. 1986.

R. Brooks. Challenges for Complete Creature Architectures. In *From Animals to Animats: Proceedings of the First International Conference on the Simulation of Adaptive Behavior*. 1990.

D. Dennett. *The Intentional Stance*. Cambridge, MA: MIT Press. 1987.

D. Dennett. *Brainchildren: Essays on Designing Minds*. Cambridge, MA: MIT Press. 1998.

R. Firby. An Investigation into Reactive Planning in Complex Domains. In *Proceedings of the Sixth National Conference on Artificial Intelligence*. 1987.

M. Johnson, A. Wilson, B. Blumberg, C. Kline, A. Bobick. Sympathetic Interfaces: Using Plush Toys to Control Autonomous Animated Characters. To appear in *Proceedings of CHI 99*. 1998.

J. LeDoux. *The Emotional Brain*. New York: Simon and Schuster. 1996.

K. Lorenz. *Foundations of Ethology*. New York: Springer-Verlag. 1973.

A. Ludlow. The Behavior of a Model Animal. *Journal of Behavior* 58. 1976.

Maes, P. 1989. “How to Do the Right Thing.” *Connection Science Journal* 1(3).

P. Maes, T. Darrell, B. Blumberg, A. Pentland. The ALIVE System: Wireless, Full-Body Interaction with Autonomous Agents. In *ACM Special Issue on Multimedia and Multisensory Virtual Worlds*. 1996.

M. Mataric. Interaction and Intelligent Behavior, Technical Report, AI-TR-1495, MIT Artificial Intelligence Laboratory. 1994.

D. McFarland. *Animal Behavior*. Harlow, UK: Longman Scientific and Technical. 1973.

M. Minsky. The Society of Mind. New York: Simon & Schuster. 1988.

C. Reynolds. Flocks, Herds, and Schools: A Distributed Behavior Model. In *Proceedings of SIGGRAPH 87*. 1987.

C. Rose, M. Cohen, B. Bodenheimer. Verbs and Adverbs: Multidimensional Motion Interpolation. *IEEE Computer Graphics & Applications* 18(5). 1998.

W.S. Reilly. Believable Social and Emotional Agents. Ph.D. thesis, School of Computer Science, Carnegie Mellon Univ. 1996.

F. Thomas, O. Johnson. *The Illusion of Life: Disney Animation*. New York: Hyperion. 1981.

X. Tu, D. Terzopolous. Artificial Fishes: Physics, Locomotion, Perception, Behavior. In *Proceedings of SIGGRAPH 94*. 1994.

T. Tyrrell. Computational Mechanisms for Action Selection. Ph.D. Thesis. Centre for Cognitive Science, University of Edinburgh. 1993.

J. Velasquez.W When Robots Weep: Emotional Memories and Decision-Making. In *Proceedings of AAAI-98*. 1998.