Automaticity and Expressive Behavior in Virtual Actors: Notes on the Organization of Mammalian Behavior Systems

Ian Horswill
Northwestern University, Evanston IL
ian@northwestern.edu

Abstract

Much of the most expressive behavior in humans - expressions of shock or alarm, gaze aversion, or explosive rage - are the result of automatic processes that engage before deliberative processing can respond. In some cases, such as weeping, the deliberative system may have only limited ability to override the automatic system. These processes are implemented by a network of phylogenetically old, special purpose, somewhat redundant systems that give rise to the particular idiosyncratic behavior we associate with automatic reactions to emotional events. In this paper, I'll review some of the ethological and neuropsychological results on low-level systems related to threat response, and their relation to the simulation of virtual characters. I will also discuss work in progress on building a medium-fidelity simulation of these systems.

Introduction

Human behavior is the result of a complex, distributed system of neural circuits, ranging from the flexible, high-level systems that underlie language and abstract reasoning, to the older, more immediately survival-oriented systems that are responsible for fast, stereotyped responses to broad classes of events such as rewards or threats. This latter class of system is responsible for much of the behavior that we see as being emotionally expressive; because the automaticity of these systems makes their behavior difficult to repress or fake, humans and other animals use them as imperfect, but generally reliable indicators of an individual's affective and attentional state. This reliance on automatic behaviors for communicating affect and other forms of internal state makes them especially important for virtual actors and for narrative more generally.

Having been evolved originally for a very different set of environmental challenges than we face today, the behavior of these systems is often unnecessary or even detrimental; it is in any case, not what a good AI system would produce given a straightforward means-ends analysis of the situation. It is instead, a peculiar product of our evolutionary heritage and the particular kinds of survival challenges faced by our distant ancestors. If someone tells you you've been laid off, the extra illumination admitted to your eyes by "shock face" (figure 1) – the raising of the brow, and widening of the lids – is of no use in finding a new job, nor will the autonomic effects designed to provide energy for the fight-or-flight system, such as increased heart rate and blood glucose.

Figure 1: shock expression
(http://www.flickr.com/photos/nicolasconnault/2974673841/, Creative Commons license)

Nevertheless, these sorts of automatic, evolutionarily idiosyncratic responses are a key part of human behavior, and ones which we expect our characters to perform. While I would not suggest that it is necessary, or even desirable, to specifically duplicate the architecture of the human brain in virtual characters, it is nevertheless useful to understand what is known about these automatic behaviors and how they are implemented in the brain.

In this paper I will discuss some general organizational structures of the mammalian brain that are interesting and relevant to expressive behavior. Then I will summarize what is known about the mediation of threat response in mammals, their relations to specific neural circuits, and their relation to human behavior and expressive behavior in

particular. Where relevant, I will discuss their relationship to contemporary theories of personality and clinical psychology. Finally, I will give a sketch of work in progress on implementing some of these structures in Sims-style interactive characters.

Organizational Structures Mammalian Behavior Systems

There are a number of aspects of the organization of the brain which are interestingly different from those of typical AI systems, and worth commenting on from the standpoint of expressive character behavior. The point here is not to argue that they are smarter (if anything, they're dumber) but rather that they are different from how you would design the system if you were starting from scratch, and that their idiosyncrasies are important to the particular ways in which human behavior is generated.

Redundancy and Temporal Hierarchy

One organizational principle that has been discussed extensively in the robotics literature (e.g. (Brooks 1986)) is the use of independent control systems operating in parallel, often arranged in rough hierarchies. Examples of control systems inspired by these notions include the subsumption architecture (Brooks 1986), the motor schemas architecture (Arkin 1998), behavior nets (Maes 1989), and the ethnologically inspired work of Blumberg (Blumberg 1996). For example, in subsumption, functionality is broken down into layers of parallel control systems, so that each layer adds new types of functionality to the lower layers. Thus, a lower level layer may implement collision avoidance, while a separate, higher layer might implement path finding, and a still higher layer would implement systematic search of an area. This architecture allows higher layers to be removed or deactivated without interfering with the functioning of lower layers.

Biological systems however, because of their complicated evolutionary derivation, may also contain surprising amounts of redundancy, where phylogenetically older systems run side by side with newer systems that solve the same problem. Since the older systems are typically faster (but dumber), they can provide a quick interim response until the newer, smarter system can kick in and override it.

The important point for the discussion at hand is that even if the older systems' responses aren't particularly helpful, as in the shock face example above, their very automaticity makes them reliable and therefore highly expressive. That is, *because* they're automatic, they act as "tells" for others about a person's affective and attentional state, even if the person immediately acts to suppress the response. Moreover, it may continue to "bleed through" the response of the higher level system, even after it kicks in and attempts to suppress it.

Control Through Excitation and Inhibition

Importantly, the coordination between subsystems is often achieved through mutual excitation and inhibition. When one system wants to override another system, it will often do so by inhibiting the activation or motor outputs of the other system. The advantages and disadvantages of this approach have been discussed extensively in the behavior-based robotics literature (Arkin 1998). For our purposes, what matters is that this pattern of control produces characteristic side effects that are important in human behavior.

For expressive characters, the most important aspect of inhibitory control is its characteristic failure modes. When someone tells you of the death of a loved one, you may well cry. This is an automatic response that you have only very limited control over. However, you may also want to put up a brave face for those around you, so you try to at least suppress those aspects over which you have voluntary control; you make your face impassive; you control your breathing; you may even be able to regulate your voice stress. But you'll have much more limited control over the tearing, and reddening of your eyes because they aren't under voluntary control. And even those aspects you can control, such as your facial expression, still won't look normal. A stiff upper lip is not a normal lip, but one of exaggerated stiffness. Inhibition of an output often leads to hypercorrection, inhibiting *all* expression, and resulting in an impassive face.

This is important because all these behaviors – uninhibited, fully inhibited, and imperfectly inhibited – are used by others to interpret a person's internal state. Actors manipulate these expressions to deliberately communicate the desired internal state of their character.

For example, it is often held to be important for protagonists in mainstream fiction to be seen to suffer, so as to create empathy in the reader or audience. But it's equally important that they suffer nobly (Card 1999). A protagonist who gives in to her suffering and cries the way a child would is seen as weak and loses the audience's empathy. On the other hand, if she endures great hardship without showing any signs of it whatsoever, she'll be seen as pathological or at least so very superhuman that, again, the audience cannot empathize with her. So writers and actors are trained to show the character struggling with herself, working to keep herself under control but always with enough showing through, either as tears or as impassivity, to show the audience the character's internal struggle for control.

Separation of Impulsive and Effortful Control

One of the strangest phenomena of human behavior is that, although the neural substrates are unclear, humans behave as if they are composed of an impulsive control system[1] that opportunistically seeks rewards and avoids punishments, together with a separate, "effortful" control system that tries to override the impulsive system when it would otherwise do something stupid (Carver, Johnson et al. 2008). The control is "effortful" because one's ability to override the impulsive system is effectively a scare and easily depleted resource; performing a task that requires willpower, such as watching a disturbing film while trying to avoid making facial expressions, makes you measurably worse at performing subsequent tasks involving willpower (e.g. forcing yourself to work on a task you find unpleasant) until some refractory period has elapsed (Schmeichel and Baumeister 2004). Indeed, even forcing yourself to make relatively choices, such as choosing between a cookie and a candy bar, can measurably reduce your ability to stay on task, or to force yourself to avoid a destructive but pleasurable behavior such as binge eating (Vohs, Baumeister et al. 2008).

The evidence for the separation of the effortful and impulsive systems lies in large part in the high variability of individuals' ability to self-regulate. Effortful control develops comparatively late in life, shows significant, stable individual differences (and so is viewed as a personality trait), and is susceptible to manipulation in individuals either through pharmacological intervention (e.g. drinking alcohol) or by forcing the individual to perform unpleasant tasks.

Why does this matter for virtual characters? One reason is that it's an important personality variable for describing characters: an impulsivity knob would be a great parameter to have for your character AI. Another reason is that the phenomenon of ego depletion (the loss of effortful control after self-regulation tasks) provides an architectural model for why people act impulsively under stress: Carrie from *Sex and the City* eats that pint of Häagen-Dazs after a bad date because during the bad date she had to work so hard at *not* showing her disappointment.

Separation of Approach and Avoidance

Finally, it's common in both neuropsychology and personality psychology to model human behavior in general, and the impulsive system in particular, as being composed of independent subsystems for approaching attractive stimuli and avoiding aversive stimuli, rather than a single unified system, such as a reactive planner (Carver, Johnson et al. 2008).

[1] Often referred to as "the Homer Simpson brain."

The evidence for this, again, has to do with the fact that the sensitivities of the two systems appear to be key parameters, both as short-term state, and as long-term stable, inheritable traits. The two most important variables in trait theories of personality – extroversion and neuroticism – are interpreted as the sensitivities of the approach and avoidance systems (Carver and White 1994). In clinical psychology, state disorders (mood disorders) are frequently modeled as over- or under-sensitivity of the approach and avoidance systems (Zinbarg and Yoon 2008).

Mammalian Defense Systems

Having discussed some of the broad organizational principles of the mammalian brain, let's look in detail at one class of behaviors that have been extensively studied. One of the most fundamental tasks for an agent is to respond to external threats. The mammalian system, while originally developed to deal with threats such as predators, is still active in humans, and still mediates threat response, even to relatively abstract threats. Mammalian threat response is controlled by a variety of systems that operate across a range of temporal scales. Although far from being fully understood, we can at least at least sketch some known subsystems that are of interest to expressive behavior.

Reflexive Systems

Figure 2: Startle response (http://www.flickr.com/photos/gtmcknight/84215363/, Creative Commons license)

The most immediate responses to potential threats are reflexive. The *startle reflex* is mediated by a circuit in the reticular formation of the brainstem (the caudal pontine reticular nucleus), and causes open loop blinking of the eyes and contraction of the neck and shoulders to protect the eyes and neck area. It receives control information from the systems thought to be involved in fear and anxiety (Davis 1992), which decreases its triggering threshold in the presence of threats. Interestingly, the same nucleus also implements the rhythmic motor patterns involved in mastication (chewing), and it is speculated that this is responsible for the grinding of teeth under stress.

Less spectacular, but more common is the *orienting response*, in which high salience events that fall below the threshold for the startle response cause an automatic

reorientation of the head toward the event in question. This reorienting happens without conscious intervention, it being much faster than the cognitive system. The orientation response also involves a temporary decrease in heart rate (bradycardia), which stills the body (Porges 1995).

Fear and anxiety also modulate metabolism by way of the autonomic nervous system. The most important externally observable aspect of this is the increase in sweating that occurs when the sympathetic nervous system accelerates metabolism. However, in particularly extreme cases, the parasympathetic system can respond by *decreasing* metabolism through bradycardia and apnea (interruption of breathing). In reptiles, this is an adaptive response that conserves resources while feigning death. However, in mammals, who have considerably higher metabolic requirements, it leads to syncope (fainting) and in very rare cases, death (Porges 1995).

Gray and McNaughton's Mammalian Defense Hierarchy

In addition to the reflexive systems, there are a number of goal-directed systems involved in the avoidance of threats. They are arranged in a rough hierarchy in which the lower levels generate quick-and-dirty responses are later superseded by the responses of higher levels as they become available.

The lowest tier of defense systems implement the fight/flight/freeze system (FFFS), whose triggering is determined by "defensive distance," the distance to the threat relative to the speeds of the threat and of the animal itself (Blanchard, Blanchard et al. 1990). For large defensive distances, the system will choose to avoid the threat through flight; fighting is triggered only by inescapable threats, that is threats whose defensive distance makes flight impossible. Defensive distances near the fight/flight threshold trigger freezing behavior: the animal becomes still and waits in hopes that the threat will move away on its own, thereby enabling flight.

Even within the FFFS, there is redundant implementation of functionality, with lower-level, dumber systems kicking in at lower defensive distances and higher level systems operating at higher defensive distances for which there is sufficient time for them to operate. For example, the flight system is implemented for short defensive distances by an undirected escape behavior implemented in the midbrain periaqueductal gray matter (PAG), which essentially produces an open-loop motion away from the threat *without collision avoidance*. Directed escape (true flight with collision avoidance) is implemented separately in the forebrain (medial hypothalamus in the forebrain). The forebrain system has higher latency so the use of both systems allows the animal

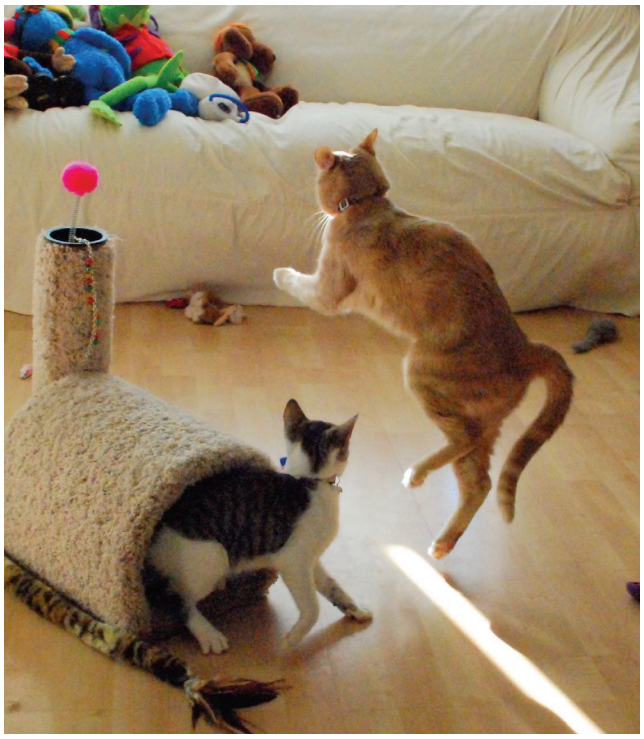

Figure 3: Undirected escape respond in domestic cat (http://www.flickr.com/photos/malingering/392390353/, Creative Commons license)

to at least start moving immediately in desperate circumstances, even if it bumps into objects in the process.

The FFFS is responsible for handling clear and present dangers. However, Gray and McNaughton argue that cases of mixed cues – situations simultaneously involving both potential threats and potential rewards – are handled by a separate system. In these "defensive approach" situations a separate Behavioral Inhibition System (BIS) responds to the conflict between approach and avoidance behaviors by inhibiting *both* systems, while triggering separate systems associated with information gathering such as visual and memory scanning in humans, and rearing in rats (Blanchard, Blanchard et al. 1990; Gray and McNaughton 2003). This produces a characteristic pattern of hesitation (separate from freezing), and slow approach, often punctuated with periodic retreats.

Finally, these systems can be overridden, to varying degrees, by higher-level cortical systems. You can decide not to run away from something, for example, although it might require a great deal of willpower. Certain other low-level behaviors however, such as drowning reflexes, are virtually impossible to override.

Role in Virtual Characters

At this point, one might understandably ask what the neuropsychology of mammalian defensive behavior has to

do with simulated humanoid characters. The canonical threat for Gray and McNaughton's theory of the mammalian defense system is after all a predator, not a house foreclosure. I've tried to outline some of the reasons above; first, to the extent that much of non-verbal communication, and expressive behavior more generally, is involuntary, it's worth understanding some of the architectural structures responsible for its being involuntary. And second, even though modern urban humans rarely face hungry tigers in the forest, their bodies respond to house foreclosures in many ways *as if they were* hungry tigers, for example, by freezing up and being unable to think.

Third, these theories have been highly influential in clinical and personality psychology of humans (Corr 2008; Fua, Revelle et al. 2010). The Gray and McNaughton model has been influential in clinical psychology, for example, because it helps explain why major depression (MD), generalized anxiety disorder (GAD) and panic disorder (PD), all of which on the face it should be modeled as overactivity of the avoidance system, are actually different disorders that respond to distinct drugs (Zinbarg and Yoon 2008). The drugs that treat panic disorder are drugs that inhibit brain areas associated with the FFFS, while the drugs that treat GAD act on brain areas involved in the BIS; depression is viewed not as overactivity in the avoidance system, but as underactivity in the approach system. So the same structures that are involved in avoiding predators in animals are theorized to be involved more generally, though mechanisms that have yet to be identified, in the wider range of human emotional behaviors.

On this argument, even if one uses an HTN planner for controlling most of a character's behavior, it's worth considering making head and gaze control independent of the planner (although the plans would presumably be able to make requests to it), because that's how it works in humans: if you're standing on the sidewalk and hear a car crash down the street, your head will have started turning to look at the crash even before your higher level systems have recognized the sound as being a crash. You may also want to think about hacking your animation system to pause any idle animations being played (e.g. to simulate breathing) because the orientation response temporarily holds the body artificially still.

Simulation

I am currently implementing a subset of these systems within Twig (Horswill 2009), a rapid prototyping system for character AI system that supports limited physical simulation and user scripting of procedural animation controllers. The goal is to be able to accurately simulate fear-related behavior, from the startle response, up to and including cognitive appraisal of events.

The undirected and directed escape systems, as well as a simple startle system, are implemented as traditional behavior-based robotics controllers that drive the animation back-end. A reimplementation of the behavioral inhibition system simulation done in collaboration with Fua, Ortony, and Revelle (2009) implements the hesitation associated with defensive approach.

The next step is to implement a simple cognitive appraisal system, mostly based on Marsella and Gratch's EMA (2006). The system will be implemented using a simple Horn clause solver based on Yield Prolog (Thompson 2010) that will be called through a job system (Gregory 2009). As events are received by the character's main AI loop, it will post jobs to the job system, which runs in a separate thread, allowing it to be decoupled from the real-time behavior systems. The results of the appraisals will then be posted back to the character through the normal game engine messaging system (Rabin 2002) and processed by the character's main update loop.

Conclusion

Much of what we think of as emotionally expressive behavior is the result of complex interactions between largely automatic processes that are not under conscious control. Indeed, it is precisely the difficulty of controlling them that makes them useful indicators of emotional state. To the extent that we want virtual actors to simulate these kinds of behaviors, it's worthwhile to look in some detail at how these behaviors are produced in the human system, and, where reasonable, to consider duplicating some of these structures in our simulations.

This is not to say we should all drop what we're doing and study neuroscience or that we would want to do detailed simulations. But it is worth experimenting with to see what kinds of interesting, emotive behavior we can produce, particularly for Sims-like (Wright 2000) emergent narrative systems in which it's acceptable for the characters to behavior almost entirely autonomously. For systems that seek a higher level of authorial control, such as Façade (Mateas and Stern 2005), where the author and/or the drama manager have greater freedom to reach in and force a character to perform whatever action best fits the desired arc of the story, this kind of deep simulation may be counter-productive. However, for those genres affording a high degree of simulation and autonomy, this is a promising source of both generativity and expressivity.

Acknowledgements

I would like to thank my colleagues Karl Fua, Andrew Ortony, and Bill Revelle for their fruitful collaboration on simulating reinforcement sensitivity theory, without which the present work would have been impossible

References

Arkin, R. (1998). Behavior-Based Robotics. Cambridge, MIT Press.

Blanchard, R. J., D. C. Blanchard, et al. (1990). "The Characterization and Modelling of Antipredator Defensive Behavior." Neuroscience & Biobehavioral Reviews **14**: 463-472.

Blumberg, B. (1996). Old Tricks, New Dogs: Ethology and Interactive Creatures. Media Lab. Cambridge, Massachusetts Institute of Technology. **Ph.D.**

Brooks, R. A. (1986). "A Robust Layered Control System For A Mobile Robot." IEEE Journal of Robotics and Automation **RA-2**(1): 14-23.

Card, O. S. (1999). Characters & Viewpoint, Writers Digest Books.

Carver, C. S., S. L. Johnson, et al. (2008). "Serotonergic Function, Two-Mode Models of Self-Regulation, and Vulnerability to Depression: What Depression Has in Common With Impulsive Aggression." Psychological Bulletin **134**(6): 912-943.

Carver, C. S. and T. L. White (1994). "Behavioral Inhibition, Behavioral Activation, and Affective Responses to

Impending Reward and Punishment: The BIS/BAS Scale." Journal of Personality and Social Psychology **67**(2): 319-333.

Corr, P. J., Ed. (2008). The Reinforcement Sensitivity Theory of Personality, Cambridge University Press.

Davis, M. (1992). The role of the amygdala in conditioned fear. The amygdala: neurobiological aspects of emotion, memory and mental dysfunction. J. P. Aggleton. New York, Wiley**:** 255–305.

Fua, K., I. Horswill, et al. (2009). Reinforcement Sensitivity Theory and Cognitive Architecture. AAAI Fall Symposium on Biologically Inspired Cognitive Architectures (AAAI Technical Report FS-09-01). Arlington, VA, AAAI Press**:** 52-54.

Fua, K., W. Revelle, et al. (2010). Modeling Personality and Individual Differences: The Approach-Avoid-Conflict Triad. 32nd Annual Meeting of the Cognitive Science Society. Portland, Oregon, Cognitive Science Society**:** 25-30.

Gray, J. A. and N. McNaughton (2003). The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System, Oxford University Press.

Gregory, J. (2009). Game Engine Architecture, AK Peters.

Horswill, I. (2009). "Lightweight Procedural Animation with Believable Physical Interactions." IEEE Transactions on Computational Intelligence, AI and Computer Games **1**(1): 39-49.

Maes, P. (1989). "How to do the right thing." Connection Science Journal **1**: 291-323.

Marsella, S. and J. Gratch (2006). EMA: A computational model of appraisal dynamics. Agent Construction and Emotions (ACE 2006). Vienna, Austria.

Mateas, M. and A. Stern (2005). Façade.

Porges, S. W. (1995). "Orienting in a defensive world: Mammalian modifications of our evolutionary heritage. A Polyvagal Theory." Psychophysiology **32**: 301-318.

Rabin, S. (2002). Enhancing a State Machine Language through Messaging. AI Game Programming Wisdom. S. Rabin, Charles River Media.

Schmeichel, B. J. and R. F. Baumeister (2004). Self-Regulatory Strength. Handbook of Self-Regulation: Research, Theory, and Applications. R. F. Baumeister and K. D. Vohs. New York, The Guilford Press**:** 84-98.

Thompson, J. (2010). "Yield Prolog." Retrieved March 15, 2011, from http://yieldprolog.sourceforge.net/.

Vohs, K. D., R. F. Baumeister, et al. (2008). "Making choices impairs subsequent self-control: A limited resource account of decision making, self-regulation, and active initiative." Journal of Personality and Social Psychology **94**: 883-898.

Wright, W. (2000). The Sims, MAXIS/Electronic Arts.

Zinbarg, R. E. and K. L. Yoon (2008). RST and Clinical Disorders: Anxiety and Depression. The Reinforcement Sensitivity Theory of Personality. P. J. Corr, Cambridge University Press**:** 360-397.