Men Are Dogs (and Women Too)

Ian Horswill

Departments of EECS and Radio/Television/Film, Northwestern University
2133 Sheridan Road, Evanston IL 60208 USA
ian@northwestern.edu

Abstract

One of the primary differences between humans and other animals is our capacity for high level cognition and the use of language. Historically, AI has understandably focused on duplicating these aspects of human behavior, and selected its architectures accordingly. However, humans are nevertheless social mammals and share a large part of the mammalian behavior repertoire, such as feeding, as well as social behaviors, such as attachment, affiliation, territoriality, and the formation of dominance hierarchies. In this paper, I will argue that for all our unique capabilities, the mammalian behavior hardware, whatever it may be is still active in humans, and that for applications such as virtual characters for interactive drama, we should begin from architectures based on our commonalities with other social mammals, rather than from our unique capabilities.

Introduction[1]

Humans are distinguished from other animals in large part by our flexible, generalized capacities for language, reasoning, and planning. While other animals may have limited capacities for simple language, sequence learning, categorization, etc., they don't share our abilities for abstract reasoning and representation. Not surprisingly, AI research has focused on flexible, general reasoning capabilities, and cognitive architectures based on some central problem-solving mechanism (Anderson, 1996; Laird, Newell, & Rosenbloom, 1987). Similarly, most mobile robots use some form of layered architecture (Bonasso, Firby, Gat, & Kortenkamp, 1997), incorporating a planner, executive, and behavior system. Similar architectures are also seen in virtual characters for interactive entertainment (Cavazza, Lugrin, Pizzi, & Charles, 2007; Mateas & Stern, 2002).

Although these systems differ significantly in details, they share the property that all or nearly all goals in the system are handled by the same subsystem(s), regardless of their type or origin. There are strong arguments for this from an engineering standpoint; for example, Newell (1990) has argued persuasively for the need for a uniform representational medium underlying reasoning system. And from the standpoint of trying to model human (domain independent) problem solving skills, this is a perfectly natural approach.

On the other hand, to the extent that other animals don't have general-purpose problem-solving abilities, this isn't a good model of how the rest of the animal kingdom works. So if we take current AI architectures too literally as models of human cognition, then we're forced to assume the animal behavior systems either somehow disappeared entirely in the transition from other apes to humans, or that, while still present, their functions have been completely subsumed by the new human faculties, whatever they may be.

In this paper, I will argue that there isn't much evidence to suggest there has been a wholesale restructuring of the higher-levels of the nervous system between humans and other social mammals. Since humans share a great deal of social behavior with other species, it's reasonable to suspect that the basic behavioral systems of social mammals are still intact in humans, running alongside the human-specific capabilities, and able to recruit those systems to accomplish their goals. To put it somewhat snarkily, humans are dogs with large forebrains.

For most AI applications, this might not be an important issue. However, I'll argue that interactive entertainment applications make these kinds of issues more relevant to applications than they might have been 20 years ago. Then, I'll present three arguments for why we might want to think the old mammalian hardware continues to be active in humans. Then I'll wave my hands furiously about what the architecture might look like, since there's been almost no work on trying to design architectures that link animal behavior with higher level cognition. Finally, I'll discuss some early efforts at simulating one particular social behavior system, the Attachment Behavior System, that's been relatively well studied.

This is, unfortunately, the kind of position paper that tends to cause the people who already agree with its conclusions to say "yes, obviously," and those who don't to say "I'm not persuaded." Although it may not change anyone's opinions, I hope it's at least useful to try to capture some version of the argument on paper.

Stories and Characters

Many have argued before that the entertainment industry is a fertile application area for AI research. Laird and van

---

Lent (2001) have specifically argued that games are AI's killer app, because they offer an application area for limited, but human-like intelligence. Bates has argued more generally (1994) for the creation of AI-based characters with rich emotional and psychological states. As Bates argued, the problems that are central to AI-based characters are not necessarily the problems that are central to more traditional application domains in AI; you don't want your blood disorder diagnosis system to get bored and stop paying attention at the end of a long work day, but you may well want a character in an interactive story to do so.

A friend of mine who is a film maker is fond of saying that films are about "blood, sweat, and tears"; that is, aggression and death (blood), sex and romance (sweat), loss and trauma (tears). We don't generally make films about chess playing. When we do, the directors and screenwriters will feel the need to add "conflict" to them, and more often than not, conflict is specifically *interpersonal* conflict. *Searching for Bobby Fischer* (Zaillian, 1993) isn't about chess so much as it's about sportsmanship. And despite its metaphysical themes, *The Matrix* (Wachowski et al., 1999), devotes several orders of magnitude more screen time to Hong Kong-style fight sequences than to Baudrillard's *Simulacra and Simulation* (1994). Robert McKee, the closest thing to a guru of Hollywood screenwriting, goes so far as to make this an ontological claim about stories in general. He argues (1997) that a story consists of a hierarchical grouping of units, each of which culminates in a reversal of "story values," such as a character's social standing or safety, or their relationship to a friend, enemy, or loved one. Higher level units produce bigger, more lasting reversals.

Although McKee argues that story should be about "universal human experience," he is actually talking in large part about universal *animal* experience. Moreover, much of popular narrative surrounds the cases in which these concerns override rational behavior.

Dogs with Large Forebrains

The argument, then, is that humans special-case a lot of processing for goals and behavior that we share with other social mammals, that we share that special-purpose hardware, whatever it may be, with other social mammals, and so for applications such as virtual characters, we're better off starting with a simulation of basic animal behavioral processes, then layering higher-level cognition on top of these, rather than by trying to implement the former within the latter. Although there isn't enough data to prove this one way or another, I'd argue that there is at least weak evidence, and it's difficult to find evidence to the contrary.

Argument 1: Phylogenetic Continuity

To begin with, it's difficult to find evidence for human neural hardware having a radically different design from other mammals. Neuroanatomical differences between species are limited, although undeniably present. The human brain is bigger than most other mammals. And there are identifiable specializations of phylogenetically older brain regions. For example, in humans the parasympathetic nervous system appears to have bifurcated into two systems, one of which appears to be specialized for certain kinds social processing (Porges, 1995). But it's difficult to find anatomical evidence for an architectural discontinuity between humans and other mammals.

Genetic differences are also surprisingly subtle. The mouse and human genomes both contain approximately 30,000 protein-coding genes, 80% of which have 1:1 orthologies with humans, meaning the corresponding mouse and human genes likely descend form a common ancestor gene, and only 1% (118) are non-homologous, meaning there appears to be no corresponding gene in the other species. Ordering is also similar; approximately 90% of the mouse and human genomes can be aligned into segments that preserve gene ordering between species (Mouse Genome Sequencing Consortium, 2002). Similar results have been found for cat, dog, and monkey.

Finally, humans share a great deal of behavior, both with other cooperative hunter species, and with other primates. One particularly well-studied example of this is the Attachment Behavior System (Bowlby, 1969), along with its less studied counterpart, the Caregiving Behavior System. These systems work together to insure that parent and child remain in sufficient proximity to insure the parent can protect the child from predators, prevent them from sticking their fingers in light sockets, and otherwise keep them out of trouble. Rhesus monkeys display nearly identical attachment behavior to young human children (Suomi, 1999).

Argument 2: Ontogenetic Continuity

The second argument for human social behavior having an independent substrate from general purpose cognition is that it's up and running long before general purpose cognition. Children are forming friendships, engaging in play with one another, forming social groups, and even bullying, long before they reach the preoperational stage of cognitive development, much less the formal operational stage. Moreover, those behaviors persist into adulthood.

For example, the Attachment Behavior System comes in long before language (Cassidy, 1999). Moreover, it persists into adulthood; both it and the caregiving system are thought to play key roles in adult romantic attachment (Feeney, 1999; Hanzan & Shaver, 1987). And dysfunctional adult attachment style is a significant risk factor for stalking behavior (Kienlen, 1998).

Argument 3: Independence of Failure Modes

The final argument for special-purpose processing of social information in humans is that it can fail independently of other modes of reasoning. Although delusional disorders are rare, they do occur. And true delusions (as opposed to memory failures, as when an Alzheimer's patient forgets a favorite restaurant has closed), are almost exclusively social. To quote a recent survey,

> "The contents [of delusions] usually relate to essential human experience and interaction: persecution, harm and jealousy …; love and sexuality …; personal greatness and glory …; guilt and failure … as well as the body …." (Kunert, Norra, & Hoff, 2007)

By and large, people don't have delusions about chess or arithmetic. The fact that the processing of information about social status, the intentions of others, or the attractiveness of our bodies can fail independently of other reasoning processes, suggests that it is processed differently than other kinds of information. Paranoia, in particular, which specifically involves delusions about threats from other agents (be they individual humans or agencies such as the government), can be triggered by schizophrenia, neural lesions, or certain types of drugs, while keeping other forms of reasoning relatively intact. In this sense, while Colby's PARRY system (1973) was a good simulation of paranoid behavior, it was not necessarily a good theory of the etiology of paranoia, because there was nothing architecturally special about these kinds of social information. PARRY could just as easily have had delusions about arithmetic or the color of someone's tie as about the mafia. (That said, PARRY did also have special-purpose processes for certain kinds of social information, such as the agent's trust level for its interrogator).

Behavioral architecture

To summarize, although there's no conclusive evidence that human cognitive architecture is the same as that of other mammals, it's even harder to find evidence that the mammalian systems have been completely replaced by a clean, new, uniform reasoning system that's responsible for all behavior and problem solving. Whatever is responsible for higher level human faculties such as language, *the old hardware is likely still up and running.*

Although the functional structure of mammalian behavior isn't understood in detail at a computational level, it is generally modeled as a collection of simpler, relatively autonomous behavioral systems running in competition, with some kind of arbitration mechanism to prevent them from generating contradictory outputs. Individual behaviors or groups of behaviors seek to achieve different goals, such as feeding, mating, nest building, *etc*. Each behavior computes what amounts to a running estimate of the behavior's utility, and the arbitration mechanism insures that the highest utility behavior is chosen subject to constraints such as a preference for completing the current goal over switching to a new one (see Blumberg (1996, chapter 4) for an accessible introduction to ethology written for an AI audience). This then suggests an architecture consisting of a set of mammalian behaviors, both basic survival behaviors, such as feeding, seeking shelter, fight, flight, and freezing, etc., and social behaviors such as group affiliation, territoriality, dominance, etc.

I won't try to take a position on how abstract thinking, language, and general problem solving are implemented architecturally. It might be a matter of older subsystems evolving (e.g. memory systems getting bigger capacity), new subsystems being introduced, or existing subsystems being connected in the right ways. It doesn't take much architecturally, to make a system Turing complete.

Relation to cognition

To me personally, the most interesting question is how the older systems interoperate with higher-level cognition, however it's implemented. Again, I'll focus on the case of the Attachment Behavior System, since it's relatively well studied. As mentioned above, young human children behave more or less indistinguishably from rhesus monkeys in most attachment experiements. That means that in both species, juveniles want to maintain proximity to the caretaker, to retreat to the caretaker when under stress, and are soothed by the caretaker's proximity, gaze, and touch. Separation from the caretaker produces high levels of anxiety.

What's important here is that the system behaves more or less like a sensory-motor system you might find in a robot: it monitors a specific set of perceptual conditions (caretaker proximity, line-of-sight, gaze, and contact); it maintains a set of internal state variables, such as anxiety; and it responds with a fairly stereotyped set of behaviors in response to perceptual and internal triggers. However, in humans, as opposed to other primates, the parental *proximity* is gradually abstracted to parental *availability*, particularly emotional availability. For example, as the child develops language, the child is able to substitute (verbal) negotiation with the parent in the form of parental reassurance that they will return, for actual proximity of the parent. Similarly, although soothing behaviors involving gaze and touch continue throughout childhood, and into adulthood, verbal soothing, and particularly talk about feelings become increasingly common.

What's interesting about this from an AI standpoint is that the attachment system behaves neither like the kinds of sensory-motor systems we commonly see in the lower layers of three-level architectures such as (Bonasso et al., 1997), nor entirely like the kinds of reasoning systems that we commonly see as the higher levels. It behaves almost like an inverted TLA, in that the "higher-level", generative system is subordinate to the "lower-level" system. In some ways, that's unsurprising; it's common to divide animal behavior into consumatory behavior and appetitive behavior. In the case of drinking, the actual drinking is a consumatory behavior; it's triggered by an internal need and extinguishes that need when it completes. However, going to the watering hole, or searching for water, are appetitive behaviors; they move the agent into a state where the consumatory behavior is possible. The smart appetitive behaviors, which involve path planning, are activated by and subordinate to the comparatively dumb drinking behavior.

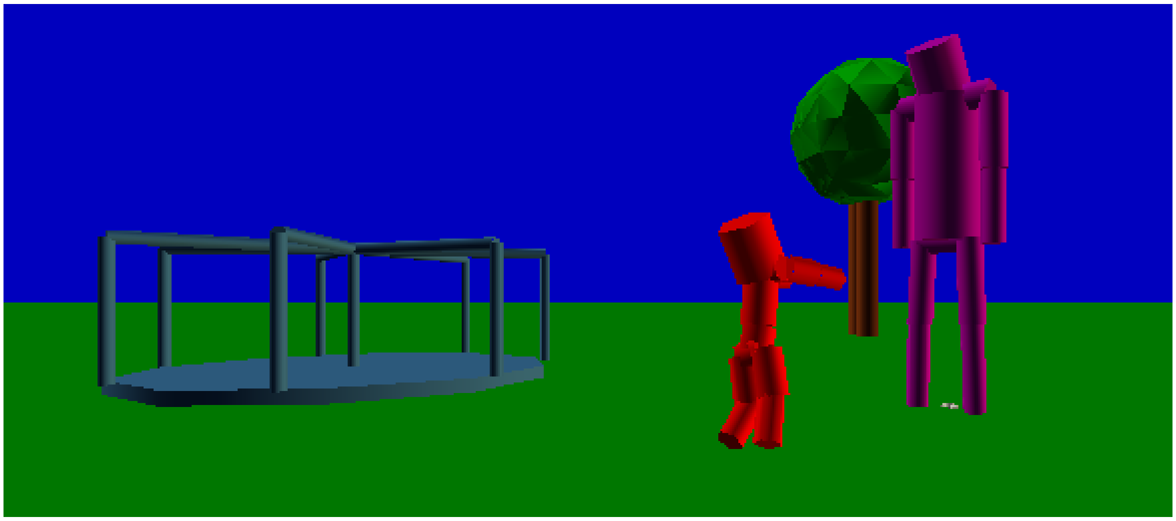

Figure 1: Simulated attachment behavior

To the extent that most of problem solving behavior is appetitive, it's can be viewed as natural for problem solving to be subordinate to the behaviors that generate the original goal.

But that's not an entirely satisfying explanation either. Because the interesting thing about attachment from an AI standpoint is that it becomes abstracted as the child develops; parental availability shifts from being a purely perceptual condition to being something more like an assessment of the probability the parent would come to the child's rescue should the need arise. If you believe the arguments I've made above, then the interesting question is: what would the rhesus monkey attachment system look like such that it could be interfaced with the kinds of higher-level systems humans have, with little or no modification, and go through the kind of developmental arc we see in human children? And, conversely, what do the high level systems look like such that they can not only act as appetitive behavior for the older systems, but also "reach in" to those systems and change their internal state variables, as when a child is soothed by language.

Steps toward implementation

If you believe the foregoing arguments, then the natural project would be to (1) implement a simulation of basic mammalian social behavior, so far as we understand it, but targeted to humanoid bodies, (2) see in what ways their behavior seemed to match or deviate from human behavior, and (3) look for natural ways of layering higher level capabilities on top of this basic substrate.

As a first step toward this, I've been working on a simulation of the "safe home base" phenomenon from Attachment Theory (Bowlby, 1969), in which a child makes excursions from the caregiver to explore the environment, but periodically returns to the caregiver to be soothed. The children display three main high-level behaviors: playing with the ball, fighting, and running to hug the parent (attachment). The characters have attention and short-term memory subsystems that appraise each object in view or in the STM for its salience (interest level), valence (hedonic value), and monitoring priority. On each update cycle, the maximal salience object becomes the focus of attention. Individual behaviors react to the focus of attention and change their activation levels accordingly.

In parallel, the gaze control system shifts visual attention between the current focus of attention, the target of the approach system (if different), and other objects that have high monitoring priority (the parent and any potential threats).

The result is that the children run after the ball because it's highly valenced, but as the small child gets farther from the parent, it becomes anxious and the monitoring priority of the parent increases, causing the child to periodically stop and look back to the parent. Eventually, the child's anxiety becomes sufficient for it to abandon the ball and return to hug the parent, which reduces the child's anxiety. Eventually, the child's attention returns to the ball, and the child returns to play.

Conclusion

I've tried to argue that not only do we share a number of basic social behaviors with dogs, monkeys, and other social mammals, but that in all likelihood we share the neural substrates that underlie them. That's not to say that social behavior could only be implemented in this manner; one could perfectly well add social reasoning and goals to a general-purpose planning and reasoning architecture. And there have been a number of systems built that work in more or less this way (Gratch & Marsella, 2001; Mateas & Stern, 2002; Reilly, 1996).

However, for characters in interactive narrative, much of what we look for are the ways in which, gaze, posture, hesitation, etc. signify the character's desires and conflicts. Again, these could be generated through explicit reasoning about affective displays (Sengers, 1998). However, an attractive option, at least for emergent narrative systems such as *The Sims* (Wright, 2000), is to simulate the processes that underlie such displays.

References

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*, 355-365.

Bates, J. (1994). The Role of Emotion in Believable Agents. *Communications of the ACM, 37*(7), 122-125.

Baudrillard, J. (1994). *Simulacra and simulation*. The Body, in theory. Ann Arbor: University of Michigan Press.

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

Bonasso, P., Firby, R. J., Gat, E., & Kortenkamp, D. (1997). Experiences with an Architecture for Intelligent Reactive Agents. *Journal of Theoretical and Experimental Artificial Intelligence*, *9*(2-3).

Bowlby, J. (1969). *Attachment and Loss*. New York,: Basic Books.

Cassidy, J. (1999). The Nature of the Child's Ties. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of Attachment* (pp. 3-20). New York: The Guilford Press.

Cavazza, M., Lugrin, J.-L., Pizzi, D., & Charles, F. (2007). *Madame Bovary on the Holodeck: Immersive Interactive Storytelling*. Augsburg, Germany: ACM Press.

Colby, K. M. (1973). Simulation of Belief Systems. In R. C. Schank & K. M. Colby (Eds.), *Computer Models of Thought and Language* (pp. 251-286). San Francisco: W.H. Freeman and Company.

Feeney, J. A. (1999). Adult Romantic Attachment and Couple Relationships. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of Attachment* (pp. 355-378). New York: Guilford Press.

Gratch, J., & Marsella, S. (2001). *Tears and Fears: Modelling Emotions and Emotional Behavior in Synthetic Agents*. Montreal: ACM Press.

Hanzan, C., & Shaver, P. R. (1987). Romantic Love Conceptualized as an Attachment Process. *Journal of Personality and Social Psychology, 52*, 511-524.

Kienlen, K. K. (1998). Developmental and Social Antecedents of Stalking. In J. R. Meloy (Ed.), *The Psychology of Stalking* (pp. 52-69). San Deigo: Academic Press.

Kunert, H. J., Norra, C., & Hoff, P. (2007). Theories of Delusional Disorders: An Update and Review. *Psychopathology, 40,* 191-202.

Laird, J. E., & Lent, M. v. (2001). Human-level AI's killer application: interactive computer games. *AI Magazine*(Summer 2001), 15-25.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An Architecture for General Intelligence. *Artificial Intelligence*, *33,* 1-65.

Mateas, M., & Stern, A. (2002). A Behavior Language for Story-Based Agents. *IEEE Intelligent Systems*, *17*(4), 39-47.

McKee, R. (1997). *Story : substance, structure, style and the principles of screenwriting* (1st ed.). New York: ReganBooks

Mouse Genome Sequencing Consortium. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*, 520-562.

Newell, A. (1990). *Unified theories of cognition*. The William James lectures. Cambridge, Mass.: Harvard University Press.

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

Reilly, S. N. (1996). *Believable Social and Emotional Agents.* Carnegie Mellon University, Pittsburgh, PA.

Sengers, P. (1998). *Do the thing right: an architecture for action-expression*. Minneapolis, MN: ACM Press.

Suomi, S. J. (1999). Attachment in Rhesus Monkeys. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of Attachment* (pp. 181-198). New York: Guilford Press.

Wachowski, A., Wachowski, L., Reeves, K., Fishburne, L., Warner Bros., Village Roadshow Pictures, et al. (1999). The Matrix [videorecording]. Burbank, Calif.: Warner Home Video,.

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

Zaillian, S. (Director). (1993). *Searching for Bobby Fischer*. United States: Paramount Pictures.