Bayesian Networks for Modeling Emotional State and Personality: Progress Report

Jack Breese **Gene Ball**
Microsoft Research
One Microsoft Way
Redmond, WA 98052-6399
breese@microsoft.com,geneb@microsoft.com

Abstract

We describe a framework for constructing a model of emotions and personality for a computational agent. The architecture uses dynamic models of emotions and personality encoded as Bayesian networks to 1) diagnose the emotions and personality of the user, and 2) generate appropriate behavior by an automated agent. Classes of interaction that are interpreted and/or generated include such things as choice of wording, characteristics of speech (speed and pitch), gesture, and facial expression. In particular, we describe the structure of dynamic Bayesian networks (DBNs) that form the basis for the interpretation and generation, and address assessment and calibration of static and dynamic components.

Introduction

Technologies such as speech recognition, text-to-speech, video input, and advances in computer graphics are providing increasingly rich tools to construct human-like user interfaces. One aspect of developing such a capability is the ability of the system to recognize the emotional state and personality of the user and respond appropriately (Picard 1995; Reeves & Nass 1995). Research has shown that users respond emotionally to their computers, and in many instances cognitive intelligence is linked to an agent's "emotional" intelligence. In order to be an effective communicant, a computer character needs to respond appropriately to a wide range of cues from the user and should produce its own emotional signals that reinforce, rather than confuse, its intended communication.

In a previous paper (Breese & Ball 1998), we described a framework using a probabilistic Bayesian network to:

- Infer or diagnose the likely emotional state and personality of the user, and
- Generate behavior in an agent (e.g. speech and gesture) consistent with a desired personality and emotional state.

In this paper, we extend that work to the domain of dynamic Bayesian networks (DBNs) (Tatman 1985; Dean & Kanazawa 1989). The temporal structure of emotional states is captured in a temporal lag structure in the dynamic Bayesian network, a generalization of the hidden Markov models posited by Picard (Picard 1995; 1997). The system is able to predict current and future emotional states based on a history of interaction (Shafer & Weyrath 1997).

Modeling Emotions and Personality

In this work, we adopt a temporal model in which current emotional state and long term personality style are characterized by discrete values along a small number of dimensions. These internal states are then treated as unobservable variables in a Bayesian network model. We construct the model based on temporal dependencies between current and previous values of variables characterizing emotions, as well as on causal relations from emotion and personality variables to observable quantities (expressions of emotion and personality) such as word choice, facial expression, speech speed, etc. The basic structure of the model is shown in Figure 1.

Our scientific understanding of emotion, personality, and their manifestations, while extensive, is far from complete. This incomplete knowledge motivates the use of Bayesian networks as an appropriate modeling tool. We adopt a probabilistic perspective for expressing this uncertainty, and Bayesian networks provide a general framework for expressing the appropriate dependencies among variables. As discussed below, Bayesian network algorithms can be used to perform causal inference (from causes to effects) as well as diagnostic reasoning (from effects to causes). These types of reasoning correspond directly to the diagnostic and generative tasks inherent in constructing affective systems. On the representation side, the flexibility of dependency structures expressible within the Bayes net framework make it possible to integrate various aspects of emotion and personality into a single model that is easily extended and modified.

Emotion is the term used in psychology to describe short-term variations in internal mental state, including both physical responses like fear, and cognitive responses like jealousy. We focus on two basic dimensions of emotional response (Lang 1995) that can usefully characterize wide range of emotional responses:

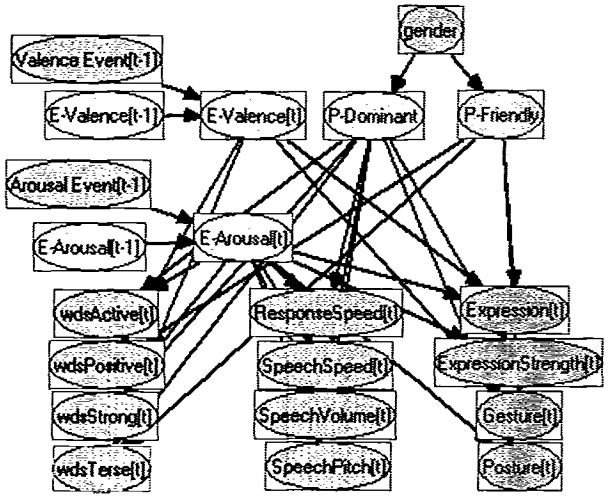

Figure 1: A Bayesian network indicating the components of emotion and personality and various types of observable effects.

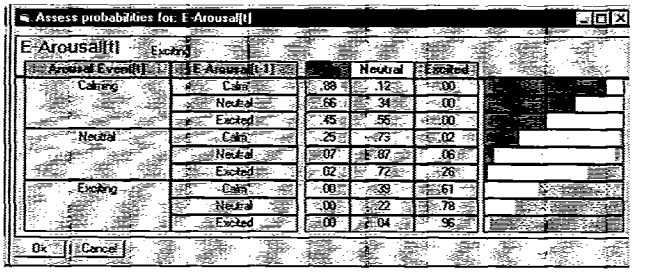

Figure 2: Probability distribution for various current levels of arousal conditioned on the previous level of arousal and the occurrence of an arousal event.

Men		Women		Total	
Mean	SD	Mean	SD	Mean	SD
5.87	0.95	5.72	1.01	5.79	0.99

Table 1: Pyschometric Characteristics of Interpersonal Adjectives for the PA-Ambitious-Dominant scale (Wiggins 1979).

- **Valence** represents overall happiness encoded as *positive* (happy), *neutral*, or *negative* (sad).
- **Arousal** represents the intensity level emotion, encoded as *excited* , *neutral*, or *calm*.

These variables evolve over time, therefore we assume a temporal process for their causal structure, as shown in Figure 1. In this network the interval between time slices is posited to be three seconds. Valence, modeled by the variable *E-Valence[t]* in the network depends on valence in the previous time period, *E-Valence[t-1]*, as well as the occurrence of *Valence Event[t-1]* in the previous period. A valence event refers to an event in the interaction that effects valence. For example in a troubleshooting application, a negative valence event might be a failed repair attempt or a misrecognized utterance. We have a similar structure for arousal, where the variable *Arousal Event[t-1]* captures external events that may effect arousal in the current period, with discrete states *Calming*, *Neutral*, and *Exciting*. The conditional probability distribution indicating the Markov transition probabilities is shown in Figure 2. The distribution does not admit a direct transition from a Calm state of arousal to an Excited state. Note this distribution is illustrative, and not based on a formal study or experiment.

Personality characterizes the long-term patterns of thought, emotion, and behavior associated with an individual. Psychologists have characterized five basic dimensions of personality, which form the basis of commonly used personality tests. We have chosen to model the two traits (McCrae & Costa 1989) that appear to be most critical to interpersonal relationships:

- **Dominance** indicates a disposition toward controlling or being controlled by others, encoded as *dominant*, *neutral*, or *submissive*.
- **Friendliness** measures the tendency to be warm and sympathetic, and is encoded as *friendly* , *neutral*, or *unfriendly*.

Since personality by definition is a long-term trait, we treat these variables as not being time dependent in the model, hence the lack of a time index in the variable names for personality, *P-Friendly* and *P-Dominant*.

The probability distributions for Dominance and Friendliness were based on a study by Wiggins (Wiggins 1979). The data comes from Table 3, Psychometric Characteristics of Interpersonal Adjective Scales. This table provides data for a sample of 610 university students. They rated themselves with respect to descriptive accuracy (on a nine point scale) over 16 interpersonal adjective scales. These scales were then reduced to 8 principal components. The means and standard deviations for men, women, and the entire sample were reported for these eight factors. Given the differences in ratings for males and females, we condition the personality variables on gender in the network.

In order to generate probabilities for the Bayesian network, we first define thresholds for a set of discrete values of each personality dimension. We then calculate the frequency that an individual will have a score in each discrete category based on the distributional information provided for men and women. The calculation for the PA (Ambitious-Dominant) dimension was performed using the reported statistics from Table 1.

We take the Total category and define "Submissive" as having a score of less than the mean minus .75 standard deviations, and "Dominant" as having a score of more than the mean plus .75 standard deviations. this

Dominance	Submissive	Neutral	Dominant
Male	.20	.56	.24
Female	.25	.54	.21
Friendliness	Unfriendly	Neutral	Friendly
Male	.35	.50	.15
Female	.16	.54	.30

Table 2: Conditional Probabilities for Dominance and Friendliness

creates a 1.5 SD window defined as "Neutral". The definition of each discrete value becomes:

- **Dominant** PA factor score of 6.53 or greater.
- **Neutral** PA factor score of between 5.05 and 6.53.
- **Submissive** PA factor score of less than 5.05.

We then consult the cumulative normal distribution function to calculate the probability that a person would have a Dominant, Neutral or Submissive score based on the gender-specific means and standard deviations reported above. The results for Dominance and Friendliness (based on the LM Warm-Agreeable scale) are shown in Table 2.

Note that the existence of the personality variables in the model induce a dependency among the observables at all times, so the model is not Markovian in the sense that future observations are conditionally independent of the past, given the current unknown emotional state. However this model can be converted to a Markovian representation for inference, and we discuss this issue a bit more below.

The Bayesian network integrates information from a variety of observable linguistic and non-linguistic behaviors as shown in Figure 1. Various classes of these observable effects of personality and emotion are grouped in the figure. In the following sections, we discuss the range of non-linguistic signals that can be accommodated by our model as well as a brief discussion of incorporating alternative linguistic expression of emotions and personality into the framework.

Non-Linguistic Expression

Humans communicate their emotional state constantly through a variety of non-verbal behaviors, ranging from explicit (and sometimes conscious) signals like smiles and frowns, to subtle (and unconscious) variations in speech rhythm or body posture. Moreover, people are correspondingly sensitive to the signals produced by others, and can frequently assess the emotional states of one another accurately even though they may be unaware of the observations that prompted their conclusions.

The range of non-linguistic behaviors that transmit information about personality and emotion is quite large. We have only begun to consider them carefully, and list here just a few of the more obvious examples. Emotional arousal affects a number of (relatively) easily observed behaviors, including speech speed and amplitude, the size and speed of gestures, and some aspects of facial expression and posture. Emotional valence is signaled most clearly by facial expression, but can also be communicated by means of the pitch contour and rhythm of speech. Dominant personalities might be expected to generate characteristic rhythms and amplitude of speech, as well as assertive postures and gestures. Friendliness will typically be demonstrated through facial expressions, speech prosody, gestures and posture. In our network we have indicated a set of nodes related to speech characteristics as well as a set related to movement, gesture, and facial expression.

The observation and classification of emotionally communicative behaviors raises many challenges, ranging from simple calibration issues (e.g. speech amplitude) to gaps in psychological understanding (e.g. the relationship between body posture and personality type). However, given an appropriate sensor (e.g. a gesture size estimator from camera input), the addition of a new source of information to the model is fairly straightforward by adding the appropriate node for the sensed input, and conditioning on relevant causative factors.

Selection of Words and Phrases

A key method of communicating emotional state is by choosing among semantically equivalent, but emotionally diverse paraphrases– for example, the difference between responding to a request with "sure thing", "yes", or "if you insist". Similarly, an individual's personality type will frequently influence their choice of phrasing, e.g.: "you should definitely" versus "perhaps you might like to".

We have modeled wording choice more deeply than other aspects of the emotion and personality. As indicated in the network in Figure 1, there are a number of "wording" nodes that correspond to the tendency to use active, positive, strong, or terse wording in making an utterance. We have devised a method for assessing individual paraphrases with respect to these dimensions, inferring the most likely paraphrase in a generative scenario, as well as interpreting recognized utterances in terms of their effect on one's belief regarding the affective state of a user. The details of the representation and inference for this aspect of the network is described elsewhere (Breese & Ball 1998).

Inference

There are two facets to the inference task using the DBN we have defined. The first is temporal reasoning regarding emotional state and personality given various pieces of evidence over time. The second addresses the generation and interpretation of word choice within the framework. We discuss each below.

Inference over Time

Dynamic Bayesian networks used to model systems that evolve over time. In their simplest form, they are equiv-

alent to hidden Markov models used in speech recognition and other temporal interpretation tasks. The particular DBN we have developed has a special structure in that we have two hidden state variables (Valence and Arousal) that are temporally dependent, and two hidden variables (Dominance and Friendliness) that are atemporal. A DBN can be *unrolled* for a given number of periods, replicating the temporal structure for each period. In Figure 3, we show the basic structure of our model replicated over three time periods. Note that the personality variables are not replicated, since they do not vary with time.

We can perform inference in this unrolled version for a fixed number of periods using standard Bayesian network inference techniques. That is, we can query the probability of future and current emotional states given a history of interaction with the agent. However, as the length of history grows, so does the size of the Bayesian network and inference becomes unmanageable. Alternatively, we can convert the network to create a single Markov hidden state that is the cross product of the 4 affective variables in the model. Here inference over time is efficient, but the size of the hidden state variable has grown exponentially.

Recently, there have been advances in utilizing the independencies in the structure of the hidden variables to render inference in DBN practical (Boyen & Koller 1998). We are investigating various methods for inference in the framework of an experimental temporal inference system we have developed.

A screen shot of this system is shown in Figure 4. The time path of the probability distribution for the 4 hypothesis variables are shown in the right 4 panes, where the size of the bar is proportional to the probability of each state. The lower left pane shows the evidential history for the scenario. In this case, we observed an exciting event in time period 1, which results in a large probability that the user is excited in period 1, and then tails off towards a steady state distribution by period 4 (12 seconds). We then recognized a wave and a happy expression in period 5. This increases the probability of excited arousal, positive valence, and that we are dealing with a friendly person. The distribution for the emotion variables tends to converge toward neutral values, while our assessment of the personality is unchanged in subsequent periods. In period 9 speech was observed to be loud, fast, and high pitched. These speech-related observations increase the probability of a dominant personality, and an excited state of arousal.

Word Choice

As discussed above, the merged Bayesian network model relates emotional state and personality to wording choice, speech characteristics, input style characteristics, and body language/movements. Most of these observable expressions are modeled as being directly caused by the components of emotion and personality. For choice of paraphrase we make an additional assumption: the individual being modeled chooses wording so as to match the intended interpretation with their current desired expressive style.

Under this interpretation, the model captures a decision model regarding word selection. The selection of a paraphrase is done such that it maximizes the probability of a match between intended expressive style and interpretation, given all previous observations regarding gesture, speech characteristics, and wording choice (Breese & Ball 1998).

Reasoning Architecture

In constructing an affectively aware agent, we maintain 2 copies of the emotion/personality model. One is used to diagnose the user, the other to generate behavior for the agent. The basic sequence of operations is shown in Figure 5. In the following, we will discuss the basic procedures used in this architecture, referring to the numbered steps in the figure.

1. Observe. This step refers to recognizing a speech characteristic, visual cue, or an utterance as one of the possible paraphrases for a concept. When such an event is recognized, the corresponding node or nodes in the user Bayesian network is set to the appropriate value.

2. Update. Here we use a probabilistic inference as illustrated in the previous section to update probabilities of personality and emotional state given the observations.

3. Agent Response. The linkage between the models is captured in the agent response component. This is the mapping from the updated probabilities of the emotional states and personality of the user to the emotional state and personality of the agent. The response component can be designed to develop an *empathetic* agent, whose mood and personality matches that of the user, or a *contrary* agent, whose emotions and personality tend to be the exact opposite of the user. Research has indicated that users prefer a computerized agent to have a personality makeup similar to their own (Reeves & Nass 1995), so by default our prototypes implement the *empathetic* response policy.

4. Propagate. Again we use a probabilistic inference algorithm to generate probability distributions over paraphrases, animations, speech characteristics, and so on, consistent with the emotional state and personality set by the response module.

5. Generate Behavior- At a given stage of the dialog, the task model may dictate that the agent express a particular concept, for example *greet* or *regret*. We then consult the agent Bayesian network for the current distribution over the possible paraphrases for expressing that concept. We can chose that paraphrase with the maximum probability, or sample from that distribution to select a particular paraphrase. This string is then passed to the text-to speech engine for generating output. Similar techniques are used to

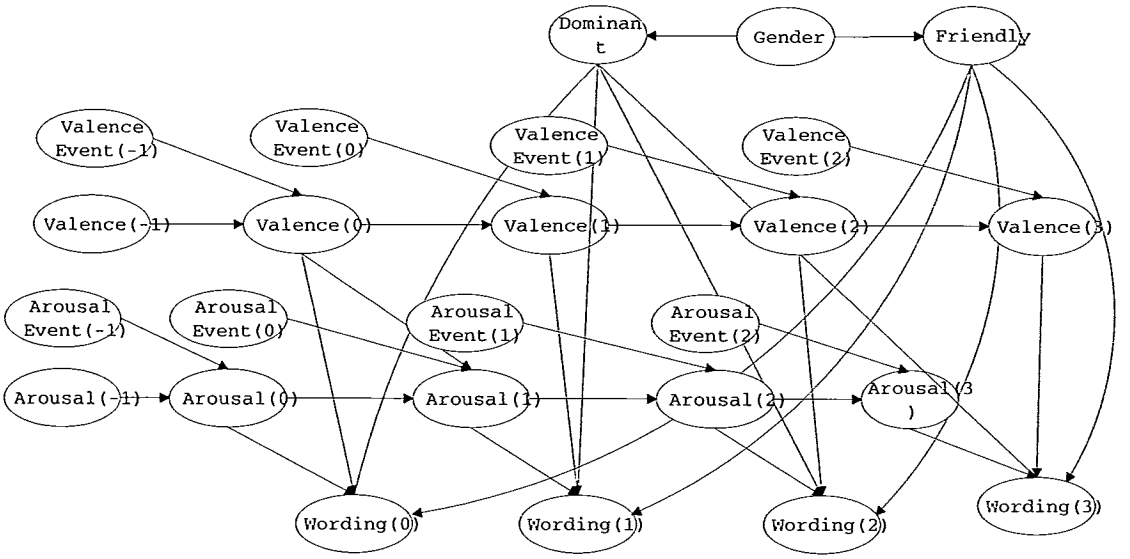

Figure 3: The unrolled version of the emotion and personality DBN for three periods.

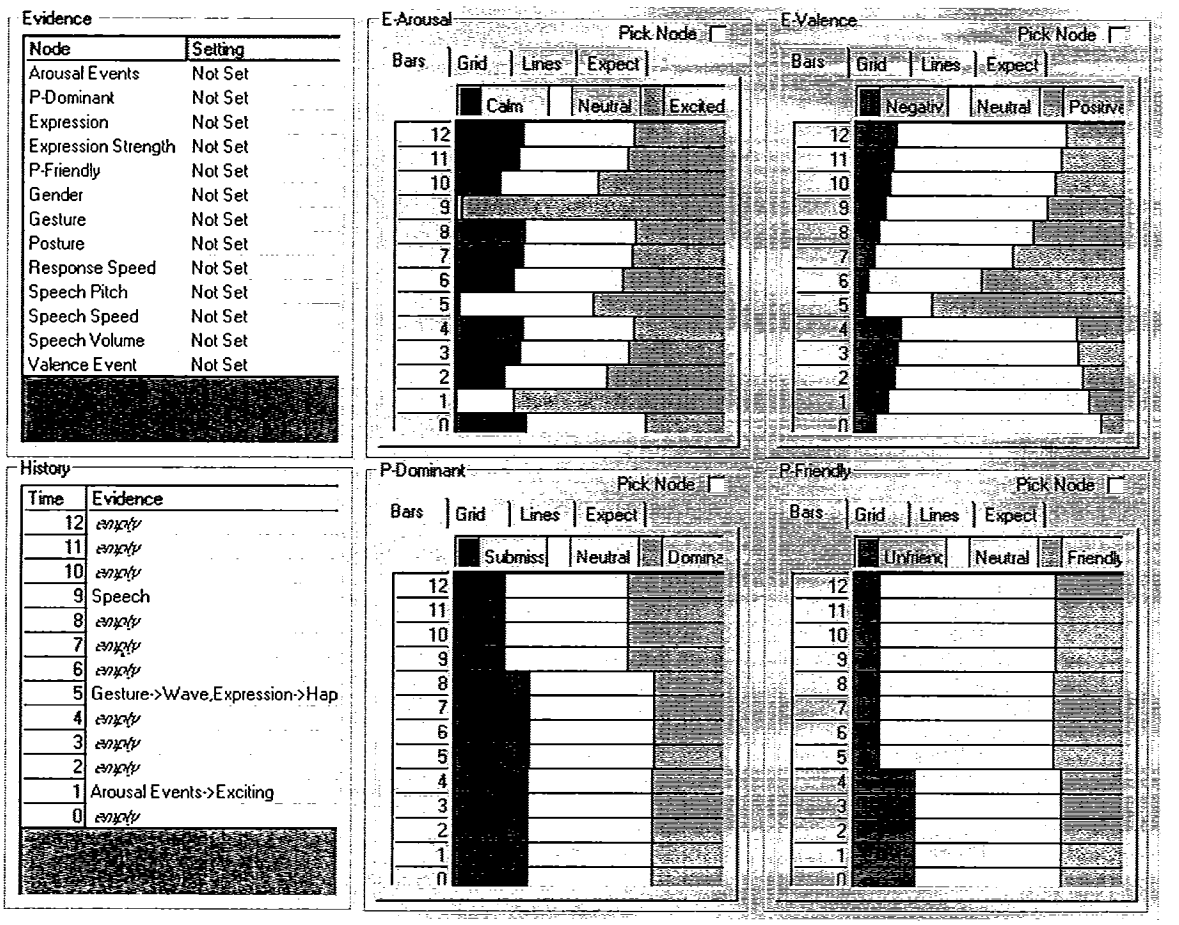

Figure 4: Display of emotion and personality probabilities for a sample scenario.

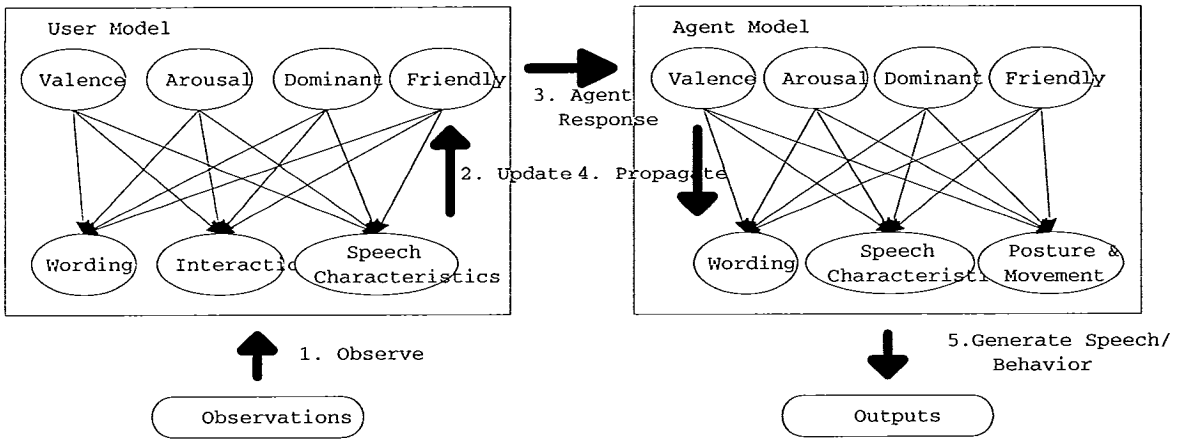

Figure 5: An architecture for speech and interaction interpretation and subsequent behavior generation by a character based agent.

generate animations, and adjust speech speed and volume, or other affect-sensitive actions.

Conclusions and Future Work

This paper presents work in progress in developing adaptive conversational user interfaces. We have presented a dynamic Bayesian network and a reasoning architecture for an agent that can recognize a user's personality and emotional state and respond appropriately in a non-deterministic manner. The model and architecture have been implemented in two prototypes at Microsoft Research. We have begun to derive assessments and validate inferences based on the psychological literature. In future studies, we plan to validate the predictions of the models by comparing them to human judgment in diagnosing emotion and personality.

Acknowledgements

We thank Clifford Nass and Byron Reeves for useful discussions and encouragement during the course of this research.

References

Boyen, X., and Koller, D. 1998. Tractable inference for complex stochastic processes. In *Proceedings of Fourteenth Conference on Uncertainty in Artificial Intelligence,* Madison, WI. Morgan Kaufmann.

Breese, J., and Ball, G. 1998. Modeling emotional state and personality for conversational agents. In *Interactive and Mixed-Initiative Decision Theoretic Systems, Papers from the 1998 AAAI Spring Symposium,* number Technical Report SS-98-03. AAAI Press. 7–13. Also appears as Microsoft Research Technical Report MSR-TR-98-41.

Dean, T., and Kanazawa, K. 1989. A model for reasoning about persistence and causation. *Computational Intelligence* 5(3).

Lang, P. 1995. The emotion probe: Studies of motivation and attention. *American Psychologist* 50(5):372–385.

McCrae, R., and Costa, P.T., J. 1989. The structure of interpersonal traits: Wiggin's circumplex and the five-factor model. *Journal of Personality and Social Psychology* 56(5):586–595.

Picard, R. W. 1995. Affective computing. Technical Report 321, M.I.T. Media Lab Perceptual Computing Section, Cambridge, MA.

Picard, R. W. 1997. *Affective Computing.* Cambridge, MA: MIT Press.

Reeves, B., and Nass, C. 1995. *The Media Equation.* New York, New York: CSLI Publications and Cambridge University Press.

Shafer, R., and Weyrath, T. 1997. Assessing temporally variable user properties with dynamic Bayesian networks. In Jameson, A.; Paris, C.; and Tasso, C., eds., *User Modeling: Proceedings of the Sixth International Conference.* New York: Springer-Verlag Wien.

Tatman, J. 1985. *Decision Processes in Influence Diagrams; Formulation and Analysis.* Ph.D. Dissertation, Department of Engineering-Economic Systems, Stanford University.

Wiggins, J. S. 1979. A psychological taxonomy of trait-descriptive terms: The interprersonal domain. *Journal of Personality and Social Psychology* 37(3):395–412.