Socially Aware Computation and Communication

By building machines that understand social signaling and social context, technologists can dramatically improve collective decision making and help keep remote users in the loop.

Alex (Sandy) Pentland
Massachusetts Institute of Technology

Wouldn't it be wonderful if people could work together more smoothly and productively? Imagine a world in which it is normal to openly speak your concerns and to have a fair and honest group discussion, in which people are enthusiastic about carrying through group decisions in a transparent and comprehensive way. Given the variety and frequency of jokes about bad meetings, and indeed about failed communication in general, such meetings and enthusiasm seem destined to remain wishful thinking.

Although developers of communication-support tools have certainly tried to create products that support group thinking, they usually do so without adequately accounting for social context, so that all too often these systems are jarring and even downright rude. In fact, most people would agree that today's communication technology seems to be at war with human society. Pagers buzz, cell phones interrupt, and e-mail begs for attention until we have to pause and wonder if we are being assimilated into some sort of unhappy Borg Collective.

Technologists have responded with interfaces that wink at us and call us by name, filters that attempt to shield us from the digital onslaught, and smart devices that organize our lives by gossiping behind our backs. The result usually feels as if the intent is to keep us isolated, wandering like a clueless extra in a computer-controlled game.

These solutions, while well meaning, ultimately fail because they ignore the core problem: Computers are socially ignorant. Researchers seem to have forgotten that people are social animals, and that their roles in human organizations define the quality of their lives. Technology must account for this by recognizing that communication is always socially situated and that discussions are not just words but part of a larger social dialogue.

This web of social interaction forms a sort of collective intelligence; it is the unspoken shared understanding that enforces the dominance hierarchy and passes judgment about whether your proposal fits with "the way things are done around here."

Successful human communicators acknowledge this collective intelligence and work with it; digital communications must begin to do the same by building tools that can accurately quantify social context and teach computers about successful social behavior.

At MIT, our research group is taking first steps toward quantifying social context in human communication. We have developed three socially aware platforms that objectively measure several aspects of social context, including nonlinguistic *social signals* measured by analyzing the person's tone of voice, facial movement, or gesture.[1]

We have found nonlinguistic social signals to be particularly powerful for analyzing and predicting human behavior, sometimes exceeding even expert human capabilities. These tools measure social context, which lets the communications system support social and organizational roles instead of viewing the individual as an isolated entity. Sample applications include automatically patching people into socially important conversations, instigating con-

0018-9162/05/$20.00 © 2005 IEEE

Prosodic style is the most revealing channel for social signals because it is the least subject to conscious control.

versations among people to build a more solid social network, and reinforcing family ties.

SOCIAL SIGNALS

Psychologists have firmly established that social signals are a powerful determinant of human behavior and speculate that they might have evolved as a way to establish hierarchy and group cohesion.[2,3]

Most culture-specific social communications are conscious, but other social signals function as a *sub*conscious collective discussion about relationships, resources, risks, and rewards. In essence, they become a subconscious "social mind" that interacts with the conscious individual mind. In many situations the nonlinguistic signals that serve as the basis for this collective social discussion are just as important as conscious content in determining human behavior.[2-5]

A mental partnership

Imagine a tribe on the African veldt. Each day the adults gather and hunt, and in the evening they return to sit around a central clearing where they recount the day's events and observations and discuss what to do tomorrow.

During the discussion, social signals, such as body posture and tone of voice, reflect the power hierarchy as well as individual desires. Each bit of new information comes with some collective social signaling that clearly communicates to each individual what the group thinks about that news or idea. By the discussion's end, the group has made many collective decisions, and the iron hand of social pressure will enforce the required individual behaviors.

Dominance displays have since given way to office politics, but the mechanism and result haven't changed much. The collective mind still uses social signals to guide individual behavior.

What are they?

Body language, facial expression, and tone of voice are some of the nonlinguistic signals that underpin this mental partnership. You might see someone taking charge of a conversation, for example, or hear a person setting the conversational tone—skills often associated with higher social status or leadership.

Others seem more adept at establishing a friendly interaction, which indicates skill at social connection, a trait many successful salespeople exhibit.[5] Prosodic style—also called tone of voice, roughly the way people vary pitch and volume in speaking—is perhaps the most powerful channel for these nonlinguistic social signals because it is the least subject to conscious control.[3]

Social psychologists have found social signals to be extremely powerful in predicting human behavior across a wide range of school, business, government, and family situations. With only a few minutes of observation, an expert psychologist can regularly predict behavioral outcome with about 70 percent accuracy.[3]

Amazingly, observing such thin slices of behavior can accurately predict important life events—divorce, student performance, and criminal conviction—even though these events might not occur until months, or sometimes years, later.

PREDICTING SOCIAL OUTCOMES

Following the social psychologists' example, a test for our ability to automatically measure social signals should also be a test of our ability to predict outcomes from observing a "thin slice" of human interactions. Could we predict human behavior without listening to words or knowing about the people involved?

Our research group has built a computer system that objectively measures a set of nonlinguistic social signals, such as engagement, mirroring, activity level, and stress, by looking at tone of voice over one-minute periods.[1] Unlike most other researchers, our goal was to measure signals of speaker attitude rather than trying to puzzle out the speaker's instantaneous internal state. Consequently, we treated prosody and gesture as a longer term motion texture rather than focusing on individual motions or accents. Although people are largely unconscious of this type of behavior, other researchers[2,3,6,7] have shown that similar measurements are predictive of infant language development, empathy judgments, attitude, and even personality development in children.

Using our social perception machine, we could listen in to the social signals within conversations, while ignoring the words themselves. We found that after a few minutes of listening, we could predict

- who would exchange business cards at a meeting;
- which couples would exchange phone numbers at a bar;
- who would come out ahead in a negotiation;
- who was a connector within a work group; and
- a range of subjective judgments, including whether or not a person felt a negotiation was

honest and fair or a conversation was interesting.

After excluding cases in which we didn't have enough signal to make a decision, our prediction accuracy averaged almost 90 percent. The "Measuring Prediction Accuracy" sidebar tells how we calculated accuracy.

Achieving this level of accuracy is pretty amazing, especially given that experiments using human judges have typically shown considerably less accuracy. Moreover, the decisions we examined are among the most important in life: finding a mate, getting a job, negotiating a salary, and finding a place in a social network. These are activities for which humans prepare intellectually and strategically for decades.

What is surprising is that the largely subconscious social signaling that occurs at the start of the interaction appears to be more predictive than either the contextual facts (attractiveness and experience) or the linguistic structure (strategy chosen, arguments employed, and so on).

Measuring Prediction Accuracy

We calculated a linear predictor of outcome by a cross-validated linear regression between the four audio social signaling features (described elsewhere) and behavioral outcome. We then compared this predictor to the actual behavioral outcome. The histogram in Figure A shows a typical case, in which the data is "Would you like to work with this person or not?

In a typical case with a three-class linear decision (yes, not enough information, no) the yes/no accuracy is almost 90 percent. Accuracy is typically around 80 percent with a two-class linear decision rule, where we make a decision for every case. More generally, linear predictors based on the measured social signals typically have a correlation of 0.65, ranging from around 0.40 to as much as 0.90.

Most experiments involved around 90 participants, typically 25 to 35 years old, with one-third being female. Recent papers and technical notes about these experiments are available at http://hd.media.mit.edu.

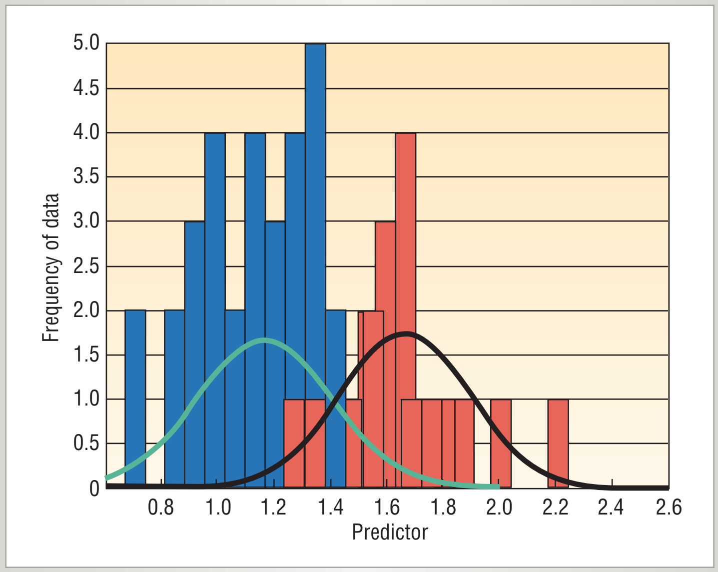

Figure A. Histogram for the data on "Would you like to work with this person or not?" The blue bars are "no" answers, the red bars are "yes." Greater predictor values mean that a "yes" is more likely. Placing the yes-no boundary at 1.4 yields a 72 percent decision accuracy

QUANTIFYING SOCIAL SIGNALS

The machine understanding community has studied human communication on many scales—phonemes, words, phrases, and dialogs, for example—and researchers have analyzed both semantic and prosodic structures. However, the sort of longer-term, multiutterance structure associated with signaling social attitude (interested, attracted, confrontational, friendly, and so on) has received little attention.

To quantify these social signals, we began by reading the voice analysis and social science literature and eventually developed texture-like measures for four types of social signaling: activity level, engagement, stress, and mirroring.[1]

By using these measurements to tap into the social signaling in face-to-face discussions, we could identify learned statistical regularities to anticipate outcomes. In addition to vocal measures of social signaling, facial and hand gesture equivalents to the audio features are being developed, and experiments using these visual features are under way

Activity level

Activity level—the simplest measure—is how much you participate in the conversation. For the activity-level measure, we use a two-level hidden Markov model (HMM) to segment the speech stream of each person into voiced and nonvoiced segments and then group the voiced segments into speaking and nonspeaking. We then measure conversational activity level by the percentage of speaking time.

Engagement

In broad terms, engagement is how involved a person is in the current interaction. Is he driving the conversation? Is she setting the tone?

We measure engagement by the influence each person's pattern of speaking versus not speaking has on the other person's pattern. Essentially, it is the measure of who drives the pattern of conversational turn taking. When two people are interacting, their individual turn-taking dynamics influence each other, which we can model as a Markov process.[6]

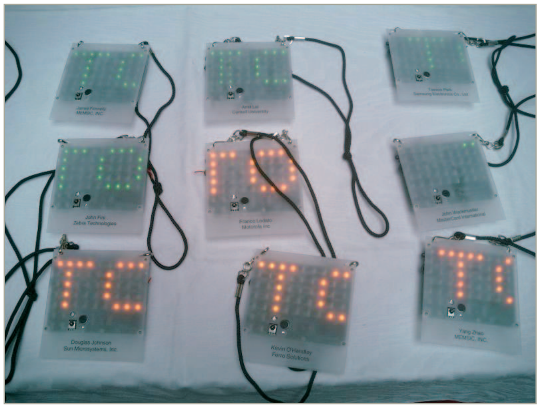

Figure 1: The Laibowitz and Paradiso Uberbadge. This badge-like platform allows social context sensing by IR, audio, and motion so that wearers can automatically bookmark interesting people and demonstrations, and it displays messages designed to build social networks.

Figure 2: The GroupMedia system. The system, built around a Sharp Zaurus PDA, measures attraction signaling in a dating and other social events. The system can also provide feedback to users and patch remote users in to interesting or socially important conversations. The interface in the image, which is still in the experimental stage, is a split screen of messages and biosignals from another user.

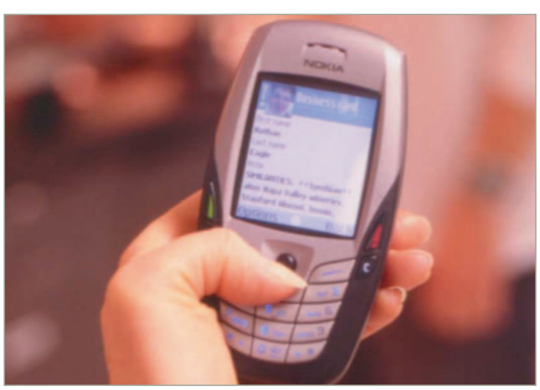

Figure 3: The Serendipity system. Built on the Nokia 6600 phone, the system senses the proximity of other people and compares their interests to make socially appropriate introductions.

By quantifying the influence each participant has on the other, we obtain a measure of their engagement. To measure these influences, we use an HMM to model their individual turn taking and measure the coupling of these two dynamic systems to estimate the influence each has on the other's turn-taking dynamics.[8]

Our method is similar to the classic work of Joseph Jaffe and colleagues,[6] who found that engagement between infant-mother dyads is predictive of language development. Our formulation relaxes those parameters so that we can calculate the direction of influence and analyze conversations involving many participants.

Stress

Stress is the variation in prosodic emphasis. For each voiced segment we extract the mean pitch (frequency of the fundamental format) and the spectral entropy. Averaging over longer periods provides estimates of the mean-scaled standard deviation of the formant frequency and spectral entropy (roughly, variation in the base frequency and frequency spread).

The sum of these standard deviations becomes a measure of speaker stress; such stress can be either purposeful (prosodic emphasis) or unintentional (caused by discomfort). Other research has used similar measures of vocal stress to detect deception and to predict the development of personality traits such as extroversion in very young children.

Mirroring

Mirroring occurs when one participant subconsciously copies another participant's prosody and gesture. Considered a signal of empathy, mirroring can positively influence the outcome of a negotiation and other interpersonal interactions.[7]

In our experiments, the distribution of utterance length is often bimodal. Sentences and sentence fragments typically occur at several-second and longer time scales. At time scales less than one second, the utterances include both short interjections ("Uh-huh.") and back-and-forth exchanges typically consisting of single words ("OK?" "OK!" "Done?" "Yup."). The frequency of these short exchanges is our measure of mirroring behavior.

INSIDE A SOCIALLY AWARE SYSTEM

We have incorporated these social signaling measurements into the development of three socially aware communications systems. Figures 1 through 3 show these systems in use. The Laibowitz and Paradiso Uberbadge is a badge-like platform,[9]

The Face of Socially Aware Communication

Four applications are being considered for commercial application:

- **Mood Ring (aka "jerk-o-meter").** Women often complain that men don't pay attention to them when they talk on the phone. The Mood Ring is a cell phone application that monitors conversations between a husband and wife and alerts the husband with a special ring tone if he is sounding inattentive or uninterested.
- **Comfort Connection.** What most of us miss when dealing with a financial institution is a friendly, trustworthy human to talk to. Comfort Connection classifies your preferred style of interaction during an initial interview, and then hooks you up with a service representative with whom you will feel comfortable working.
- **Personal Trainer.** One of the problems with the subconscious nature of social signals is that we are often unaware of how we sound to others. Consequently we often fail to put our best foot forward, most commonly when we are confused or stressed—exactly when it matters most. The Personal Trainer runs on a mobile telephone application and provides feedback at the end of each telephone call about how you sounded: aggressive, friendly, interested, firm, or cooperative. This feedback is valuable in helping you learn to present yourself in the manner you intend.
- **Winning Combination.** Businesses depend on buying low and selling high, and in most businesses this means having purchase agents and sales agents that come out ahead of the competition. The difficulty is that the style of a particular agent is not optimal for all clients—for certain pairings the company agent will tend to come out ahead, for other pairings that person will tend to lose. Winning Combination classifies the speaking style of each purchase or sales agent, and makes sure that the agent is paired with the right client.

GroupMedia[10] is based on the Sharp Zaurus PDA, and Serendipity[11] is based on the Nokia 6600 mobile telephone.

In each system, the basic element of social context is the identity of people in the user's immediate presence. The systems use several methods to determine identity, including Bluetooth-based proximity detection, infrared (IR) or radio-frequency (RF) tags, and vocal analysis.

To this basic context, it is possible to add audio feature analysis, sensors for head and body movement, and even biosignals, such as galvanic skin response (GSR). These sensing capabilities provide a quantitative measure of social context for the user's immediate, face-to-face situation. The result is a lightweight, unobtrusive, wearable system that can identify face-to-face interactions, capture collective social information, extract meaningful group descriptors, and transmit the group context to remote group members.

When the system detects a face-to-face interaction, defined as the combination of proximity and conversational turn taking, it specifies a group context that consists of the participants' identities, the four social signals, , and the compressed audio (and possibly video) information stream.

The system then creates a *social gateway* that contains the group context information and lets preapproved members of the social or work group access the ongoing conversation and group context information. The social gateway uses real-time machine learning methods to identify relevant group context changes. A distance-separated user can then access these changes.

A NEW LEVEL OF COMMUNICATION

Enabling machines to know social context will enhance many forms of socially aware communication. A simple use of social context is to provide people with feedback on their own interactions. Did you sound forceful during a negotiation? Did you sound interested when you were talking to your spouse? Did you sound like a good team member during the teleconference? Such feedback can potentially head off many unnecessary problems, as the "The Face of Socially Aware Communication" sidebar describes.

The same sort of analysis can also be useful for robots and voice interfaces. Although word selection and dialog strategy are important to a successful human-machine interaction, our experiments and those of others show that social signaling could be even more important.

What was that name?

An obvious use of social context is to help build social networks. At some time, nearly everyone has met an interesting person and then has lost that person's business card or forgotten that person's name. On the basis of an audio analysis and observations of body motion, our Uberbadge-based system[9] can keep track of all interactions during which you seem interested in the other person and e-mail you the names and particulars of those individuals at the end of the day.

Building social capital

Social capital is the ability to leverage your social network by knowing who knows what and knowing to whom you should speak to get things done. It is perhaps the central social skill for any entrepreneurial effort, yet many people find it difficult. We are therefore building systems that can help a person build social capital.

One example is the Serendipity[11] system, which is implemented on Bluetooth-enabled mobile phones and built on BlueAware, an application that scans for other Bluetooth devices in the user's proximity. When Serendipity discovers a new device nearby, it automatically sends a message to a social gateway server with the discovered device's ID. If it

Improving group function requires the ability to monitor the social communication and provide real-time intervention.

finds a match, it sends a customized picture message to each user, introducing them to one another.

The real power of this system is that it can be used to create, verify, and better characterize relationships in online social network systems, such as Friendster or Orkut.

If two people hang out after work, they are probably social friends. If they meet only at work or not at all, they are likely to have a very different relationship. The system can refine the relationship characterization by analyzing the social signaling that occurs during phone calls between the two people. The phone extracts the social signaling features as a background process so that it can provide feedback to the user about how that person sounded and to build a profile of the inter- actions the user had with the other person.

Staying in the loop

A major problem with distributed workgroups is keeping yourself in the loop. Socially mediated communications, such as GroupMedia,[10] can help with this problem by patching people into important conversations. When it detects a potentially interesting conversation, the system notifies a distant group member. Whether or not a certain member receives notification depends on measured interest levels, direction of information flow, and group membership.

A distant group member who receives a notification has several options. These options include subscribing to the information and receiving the raw audio signal plus annotations of the social context, receiving a notification from the system only in case of especially interesting comments, or storing the audio signal with social annotations for later review.

Suppose, for example, most of your workgroup has gathered, the information flow is from the boss, and the interest level is high. You might be wise to patch into the audio and track the measured level of group interest for each participant's comments. The group context information and the linking-in notification that the system gateway provides can increase both the group cohesion and your understanding of the raw audio.

The same framework could also enhance the social life of close friends. Suppose two or three of your closest friends have discovered an amazing band at a bar and are having a great time. The system could detect the situation and, given appropriate prior authorization, automatically send you an invitation to join your friends. Although such a system wouldn't be to everyone's taste, this idea generally gets a thumbs-up from college undergraduates.

Group dynamics

Social scientists have carefully studied how groups of people make decisions and the role of social context in that process. Unfortunately, they have found that socially mediated decision-making has some serious problems, including group polarization, group think, and several other types of irrational behaviors that consistently undermine group decision-making.[2,4] Improving group function requires the ability to monitor the social communication and provide real-time intervention. Human experts—facilitators or moderators—can do that effectively, but to date machines have been blind to the social signals that are such an important part of a human group's function.

The challenge, then, is how to make a computer recognize social signaling patterns. In salary negotiations, for example, we found that lower status individuals do better when showing more mirroring, which communicates that they are team players. In a potential dating situation, the key variable was the female's activity level, which indicated interest. By knowing that certain signaling patterns reliably lead to these desired states, the computer can begin to gently guide the conversation to a happy ending by providing timely feedback.

Similarly, the ability to measure social variables like interest and trust ought to enable more productive discussions, while the ability to measure social competition offers the possibility of reducing problems like group think and polarization. If the computer can measure the early signs of problems, it can intervene before the situation becomes unsalvageable.

To explore these ideas, every student in my Digital Anthropology seminar used a GroupMedia system so that our team could analyze the group interaction.[12] Real-time displays of participant interaction could be generated and publicly displayed to reflect the roles and dyadic relationships within a class. In Figure 4, the advisees (s2, s7, s8) have a high probability of conceding the floor to their professor (s9).

This type of analysis can help develop a deeper discussion. Comments that give rise to wide variations in individual reaction can cause the discussion to focus on the reason for the disparity, and those interested can retrieve these controversial topics for further analysis and debate later. The analysis also permits the clustering of opinions and comments using collaborative filtering. In this way, people can readily see opinion groupings, which sets the stage for inter- and intragroup debates.

Personal relationships

Social awareness may also be able to help reinforce family ties, an important capability in this age of constant mobility. Sensing when family members have had an unusually good, or unusually bad, experience can promote supportive communication between them.

In one version, the system would randomly leave phone messages reminding family members to call each other. However, when it senses that there has been an unusual experience—a serious argument, an especially fun conversation, or an unusually intense meeting—the system would leave reminders for others to call. The system would not tell people exactly why they should call, because doing so could violate people's privacy. Instead, the reminders would strengthen the family network by encouraging conversations precisely when family members are most likely to appreciate them.

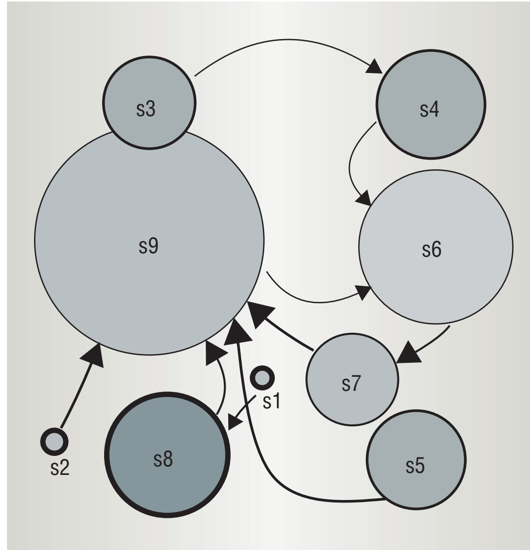

Figure 4. Display of group dynamics between professor (s9) and students during an experiment to study how a group functions. Each student in the seminar received a GroupMedia system, which analyzed the class member's interactions on the basis of activity level, group interest, and turn-taking patterns. Circle size reflects speaking time; the width of the link lines reflects the probability that the person will concede the floor. The shading within a circle reflects that person's interest level, the darker the shading, the higher the interest. Thicker circle borders denote groups.

Social signaling seems to provide an independent channel of communication, one that is quantifiable and can provide an important new dimension of communication support.

The implications of a system that can measure social context are staggering for a mobile, geographically dispersed society. Propagating social context could transform distance learning, for example, letting users become better integrated into ongoing projects and discussions, and thus improving social interaction, teamwork, and social networking. Teleconferencing might become more reflective of actual human contact, since participants could quantify the communication's value. Automatic help desks might be able to abandon their robotic, information-only delivery or their inappropriately cheerful replies.

Our current systems are just a first step toward generally useful communications tools. We must increase the reliability of our social context measurements and learn how to better use them to modulate communication. Much of our ongoing research is focusing on building meaningful mathematical models for estimating social variables and experimentally validating their use in a distance collaboration framework.

Considering the personal and societal effects of socially aware communications systems brings to mind Marshall McLuhan's "the medium is the message." By designing systems that are aware of human social signaling, and that adapt themselves to human social context, we may be able to remove the medium's message and replace it with the traditional messaging of face-to-face communication.

Just as computers are disappearing into clothing and walls, the otherness of communications technology might disappear as well, leaving us with organizations that are not only more efficient, but that also better balance our formal, informal, and personal lives. Assimilation into the Borg Collective might be inevitable, but we can still make it a more human place to live. ■

Acknowledgments

I thank my collaborators—Joost Bonsen, Jared Curhan, David Lazar, Carl Marci, M.C. Martin, and Joe Paradiso—and my current and former students—Sumit Basu, Ron Caneel, Tanzeem Choudhury, Wen Dong, Nathan Eagle, Jon Gips, Anmol Madan, and Mike Sung—for all the hard work and creativity they have added to this project. Thanks also to Deb Roy, Judith Donath, Roz Picard, and Tracy Heibeck for insightful comments and feedback. Parts of this article have appeared on Edge.org and in *Proc. IEEE Int'l Conf. Developmental Learning*.

References

1. A. Pentland, "Social Dynamics: Signals and Behavior," *Proc. Int'l Conf. Developmental Learning*, IEEE Press, 2004; http://hd.media.mit.edu.
2. C. Nass and S. Brave, *Voice Activated: How People Are Wired for Speech and How Computers Will Speak with Us*, MIT Press, 2004.
3. N. Ambady and R. Rosenthal, "Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis," *Psychological Bull.*, vol. 111, no. 2, 1992, pp. 256-274.
4. R. Brown, *Group Polarization in Social Psychology*, 2nd ed., Free Press, 1986.
5. M. Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference*, Little Brown, 2000.
6. J. Jaffe et al., "Rhythms of Dialogue in Early Infancy," *Monographs of the Soc. for Research in Child Development*, vol. 66, no. 2, 2001.
7. T. Chartrand and J. Bargh, "The Chameleon Effect: The Perception-Behavior Link and Social Interaction," *J. Personality and Social Psychology*, vol. 76, no. 6, 1999, pp. 893-910.
8. T. Choudhury, "Sensing and Modeling Human Networks," PhD dissertation, Dept. Media Arts and Sciences, MIT, 2003; http://hd.media.mit.edu.
9. M. Laibowitz and J. Paradiso, "The UberBadge Project," 2004; http://www.media.mit.edu/resenv/projects.html.
10. A. Madan, R. Caneel, and A. Pentland, "GroupMedia: Distributed Multimodal Interfaces," 2004; http://hd.media.mit.edu.
11. N. Eagle and A. Pentland, "Social Serendipity: Proximity Sensing and Cueing," 2004; http://hd.media.mit.edu.
12. N. Eagle and A. Pentland, "Social Network Computing," LNCS 2864, Springer-Verlag, 1999, pp. 289-296; http://hd.media.mit.edu.

***Alex (Sandy) Pentland** is a Toshiba Professor of Media Arts and Sciences at MIT and cofounder of MIT Media Laboratory's Digital Nations consortium, the Media Lab Asia in India, the LINCOS project in Costa Rica, and the Center for Future Health. His work encompasses wearable computing, communications technology for developing countries, human-machine interfaces, artificial intelligence, and machine perception. Pentland is a cofounder of the IEEE Computer Society's Wearable Information Systems Technical Committee, and he has received numerous awards in the arts, engineering, and sciences. Contact him at pentland@media.mit.edu.*