January 2002

# Rules for Responsive Robots: Using Human Interactions to Build Virtual Interactions

Joseph N. Cappella
*University of Pennsylvania*, jcappella@asc.upenn.edu

Catherine Pelachaud
*University of Pennsylvania*

# Rules for Responsive Robots: Using Human Interactions to Build Virtual Interactions

**Abstract**
Computers seem to be everywhere and to be able to do almost anything. Automobiles have Global Positioning Systems to give advice about travel routes and destinations. Virtual classrooms supplement and sometimes replace face-to-face classroom experiences with web-based systems (such as Blackboard) that allow postings, virtual discussion sections with virtual whiteboards, as well as continuous access to course documents, outlines, and the like. Various forms of "bots" search for information about intestinal diseases, plan airline reservations to Tucson, and inform us of the release of new movies that might fit our cinematic preferences. Instead of talking to the agent at AAA, the professor, the librarian, the travel agent, or the cinema-file two doors down, we are interacting with electronic social agents. Some entrepreneurs are even trying to create toys that are sufficiently responsive to engender emotional attachments between the toy and its owner.

**Comments**
Postprint version.

# Rules for Responsive Robots:
# Using Human Interactions to Build Virtual Interactions

By

Joseph N. Cappella
and
Catherine Pelachaud

Chapter prepared for Reis, Fitzpatrick, and Vangelisti (Eds.),
*Stability and Change in Relationships*

Joseph N. Cappella may be reached at the Annenberg School for Communication, University of Pennsylvania, 3620 Walnut St., Philadelphia, PA 19104-6220; Fax: 215-898-2024; Tel: 215-898-7059; JCAPPELLA@ASC.UPENN.EDU

Catherine Pelachaud may be reached at Università di Roma "La Sapienza", Dipartimento di Informatica e Sistemistica, Via Buonarroti, 12, 00185 Roma Italy

Computers seem to be everywhere and to be able to do almost anything. Automobiles have Global Positioning Systems to give advice about travel routes and destinations. Virtual classrooms supplement and sometimes replace face-to-face classroom experiences with web-based systems (such as Blackboard) that allow postings, virtual discussion sections with virtual whiteboards, as well as continuous access to course documents, outlines, and the like. Various forms of "bots" search for information about intestinal diseases, plan airline reservations to Tucson, and inform us of the release of new movies that might fit our cinematic preferences. Instead of talking to the agent at AAA, the professor, the librarian, the travel agent, or the cinema-file two doors down, we are interacting with electronic social agents. Some entrepreneurs are even trying to create toys that are sufficiently responsive to engender emotional attachments between the toy and its owner.

These trends are seen by some as the leading edge of a broader phenomenon – not just interactive computer agents but emotionally responsive computers and emotionally responsive virtual agents. Nicholas Negroponte answers the obvious question: "Absurd? Not really. Without the ability to recognize a person's emotional state, computers will remain at the most trivial levels of endeavor. … What you remember most about an influential teacher is her compassion and enthusiasm, not the rigors of grammar or science." (Negroponte, 1996, p. 184) The editors of *PC Magazine* do not consider emotionally responsive computers science fiction. "[I]n the not so distant future, your computer may know exactly how you feel" (PC Magazine, 1999, p. 9). Researchers at Microsoft are developing lifelike avatars to represent their owners and who could participate in a virtual meeting while the owner remains at the office available only remotely (Miller, 1999, p. 113).

Computer gurus are not the only people predicting the "emotionalization" of the human-computer interface. Scholars, such as Rosiland Picard (1997), have given serious attention to the possibility and value of programming computers and computer agents to be responsive emotionally. Part of her interest in this possibility is based on how people typically respond to computers.

Reeves and Nass (1996) have built a strong case for the "media equation," namely that people treat computers and new media like real people. Their claim is that people are primarily social beings ready to default to social judgements and evaluations even when they are dealing with inanimate entities such as computers. For example, in one of their studies people were led to believe that they were evaluating a teaching program run by one computer. When asked by the computer that had taught them how effective the teaching program was, participants offered more positive assessments than when the same evaluation of the teaching computer was asked by a different computer.

The authors argue that this result is explained by a norm of social politeness. Just as a person might direct less criticism to their own (human) teacher but direct harsher criticism toward the teacher when asked by a third party, so they did with the computer stations. The social rule of politeness was adopted as the default even when acting in a nonsocial context. In a different study, computers employing a dominant verbal style of interaction were preferred by users who possessed a dominant personality while those with submissive personalities preferred computers with a submissive style. This pattern parallels the social preferences that people have for other humans. Across a wide variety of studies, Reeves and Nass have shown that people are first and foremost social in their interactions, even when those interactions are with inanimate media rather than flesh and blood homo sapiens.

Picard reasons that if people are social even in non-social interactions, then human users should prefer to interact with computers and their representative systems that are more rather than less human. To be social and to be human is in part to be emotionally responsive. Picard's treatment of emotionally responsive computers involves reviewing literature on human emotional expression and recognition as well as recent thinking on emotional intelligence (Gardner, 1983, 1993; Goleman, 1995). She reports recent advances in automatic recognition of emotion and in work on the animation of facial displays of emotion.

The automated recognition and expression of emotion present immense problems for

programmers. However, even if these problems are solved, a large gap will remain. Affective interaction in human-computer interchanges cannot be reduced to sequences of recognition and expression. The fundamental feature of human interaction is contingent responsiveness which is not reducible to a mere sequence of recognition and expression by two agents. This chapter is about what it means to act in a way that is contingently responsive.

Our argument is essentially that modeling social interaction as it is experienced by humans requires certain mechanisms or rules without which simulated interactions are little more than the juxtaposition of two monologues.

We present our position by (1) defining responsiveness; (2) discussing computer simulation tools; (3) presenting empirical models of two person interactions; (4) describing the importance of responsive and unresponsive interactions to people; and (5) concluding with general rules for realistic virtual interaction between human and non-human agents.

## Virtual Interactions and Human Relationships

Before taking up these issues, it is fair to ask what this chapter has to do with human relationships. The development of computer simulations of human interactions is well underway. Service industries that provide simple transactions such as banking exchanges, fast food services, and so on are anxious to replace their service personnel with autonomous agents who will be the friendly, responsive representatives of the company that their more expensive, late, and sometimes surly and uncivil human counterparts are not. However, the models for such simulations – if they are to be accepted as viable replacements for humans – must have human social abilities.

Much of what is known about human social interaction is ignored by computer modelers. Instead, they often import their own assumptions into their models. Attend even one computer conference on "real characters" and you will find fascinating models, elegantly presented, but with little empirical foundation. Understanding the human and empirical basis for social interaction is crucial for AI

specialists. The science of relationships – especially human interaction in relationships – needs to be imported into the science of modeling interactions.

But does modeling virtual relationships have anything to do with understanding human relationships? The answer is an unequivocal "Yes!" in at least two senses. First, to provide useful information to computer simulators requires very precise claims and a very solid empirical base. This is a challenge to researchers who study human relationships. Our work will have little influence unless it is precise and empirically well founded.

In *Zen and the Art of Motorcycle Maintenance*, Robert Pirsig explores the differences between classical and romantic conceptions of knowing. Complex devices, such as motorcycles, can be appreciated for the beauty of their superficial structure and function or for their underlying causal operation. The latter, classical view, leads Pirsig's hero on an intellectual journey exploring what it can mean to know the underlying, unobserved structure and function of physical and social systems. He concludes that deep knowledge is knowledge that allows one to build a replica of the system being scrutinized. So it is with models of human interaction -- deep understanding comes when research and theory allow the simulation of the behaviors being modeled. The data we present on responsiveness in human interaction is pertinent to both the principles that will guide the simulations of virtual human interaction and to the parameters needed to tune the simulations.

Second, and this may sound truly strange, interactions between virtual agents or between virtual agents and human agents are a new form of relationship. Although this claim may sound like science fiction, it represents a future not far removed. What form such mediated interactions take and what implications they might have for the human agents behind them are a matter for speculation. However, their reality will depend on their programming which in turn will depend in part on the assumptions imported to the model. Successful virtual interactions between agents require realistic assumptions about the nature of human interactions. The study of virtual interactions, then, may provide insights into

human relationships in the same way that studying the successes and failures of any model of any system can provide insight into the function and design of the focal system. We may find ourselves studying virtual interactions to learn about human interactions.

**Defining Human Interaction**

The defining feature of human social interaction is responsiveness. What does it mean to be responsive? Responsiveness is not simply the generation or recognition of social signals. Nor is it just receiving and sending such signals. Neither can responsive interaction be reduced to the interleaving of two monologues, as if responsive interaction could be created from the behavior of two separate individuals juxtaposed. Responsive interactions are the regularized patterns of messages from one person that influence the messages sent in turn by the other over and above what they would otherwise be (Cappella, 1994). On this view, my rude remark to you during cocktails is not an interaction. Rather it is just a rude remark. But when my rude remark is followed by your sarcastic reply and, then, my biting insult, we have been responsive to one another, if not very polite.

Davis has defined responsive social interaction in terms of two kinds of contingency (Davis & Perkowitz, 1979). The first refers to the probability of a person's response to the actions of a partner in an interaction. The second concerns the proportion of responses related to the content of the previous message. The authors have been able to show that both of these measures of responsiveness are related to attraction to responsive others and to feelings of acquaintance. Responsiveness has been applied to physical pleasure and to verbal reinforcements as well ((Davis & Martin, 1978; Davis & Holtgraves, 1984).

Our definition of responsiveness is a conceptual relative of Davis' but more narrowly focused. Consider a conversation between two persons, A and B. Let the behavioral repertoire of person A be denoted by the set $X = (X_1, X_2, \ldots, X_N)$, where the values X, are the N discrete behaviors that can be enacted by person A at discrete intervals of time. No real loss of generality is entailed by assuming that

the behaviors are discrete rather than continuous or measured on a clock base rather than event time. Let the behavioral repertoire of person B be denoted by the set (Y) identical to the set X for A. Responsiveness is defined by two features of the contingent probability between the set of behaviors (X) and the set (Y):

eq. (1): $P[X_i(t + 1) \mid Y_i(t)] > 0$

eq. (2): $P[X_i(t + 1) \mid Y_j(t)] >$ or $< P[X_j(t + 1)]$

for at least some combination of the behaviors I and J. In words, equations 1 and 2 mean that B's behavior (the jth one, in fact) must influence the probability of A's behavior (the i th behavior) at some significant level and, more importantly, that the size of the probability must be greater than the probability that A will emit the behavior in the absence of B's prior behavior [2]. These two features insure that A's response level in the presence of B's behavior is above A's normal baseline behavior. A similar pair of equations can be written for A's influence on B. Together they constitute the necessary and sufficient conditions for mutual responsiveness.

Much of the research in modeling human interaction has been given over to coordinating components of a single person's expression. For example, generating a hostile remark requires coordination among semantic, vocal, gestural, and visual systems. Even simple matters such as head movements when improperly timed with bursts of speech can produce an odd appearance. The problems of modeling a realistic expression require attention to a range of physical systems and detailed knowledge about their interplay. The same is true for recognition systems. These individually based processes present enormous technical and theoretical problems that must be solved before realistic interactions can be built. But solving these individual problems will not solve the problem of realistic

social interaction by themselves. Realistic interaction requires modeling agents who are mutually responsive.

Our central claim in this paper is that building virtual humans capable of engaging in social interaction requires building responsive humans. What a "responsive virtual human" might be requires understanding what a "responsive human" is. To investigate this question we will proceed as follows:

1. Review literature on modeling human interaction as practiced in artificial intelligence.
2. Present data on human responsiveness showing that
    a. pairs of people in interaction cannot be constructed from the predispositions of individuals.
    b. being responsive depends on reacting contingently and appropriately to the behavior of others.
    c. being responsive requires sensitivity to the context of contingent responses.
    d. being responsive is the sine qua non of human interaction, but the degree and magnitude of responsiveness is highly variable.
3. People are sensitive to responsiveness in others (although they deny it) and that they are specifically sensitive to how emotionally responsive and polite people are to one another.

**Modeling Virtual Interaction in Artificial Intelligence**

In this section, our goal is to sketch a few of the tools employed in simulations of virtual interactions. By "agent" we mean a robot or human. The techniques of artificial intelligence and the methods of cognitive science provide the tools to build virtual humans with interactive capacity. However, the data, the rules, and the theory upon which modeling occurs must come from the study of human interaction.

Tools for Simulating the Behavior of Agents

Structure. Different levels of information are needed to describe and manipulate an agent. One level describes the structure of an agent. For example, an agent can be a set of joints and limbs. These settings are simple for a single legged robot, but much more complex for a human agent.

Procedure. The next level corresponds to procedures acting directly on the jointed figures. These procedures are used to build complex motions (Zhao & Badler, 1994). For example, to animate Marilyn Monroe and Humphrey Bogart, Magnenat-Thalmann and Thalmann (1987) used abstractions of muscle actions. They worked on specific regions, almost all of which corresponded to a single muscle.

Function. Walking (Ko, 1994), grasping an object (Rijpkema & Girard, 1991), keeping one's balance (Phillips & Badler, 1991) or expressing a facial emotion (Lee, Terzopoulos, & Waters, 1995), are very difficult to simulate if one has to work at the level of joint movements or of their equations of motion. Instead, such behaviors can be built up as functions from the lower levels of description. For example, facial animation is simulated by integrating the representation of the various layers of the facial tissue with dynamic simulation of the muscle movement (Lee, Terzopolous, & Waters, 1995). The skin is constructed from a lattice whose points are connected by springs. To carry out an animation the user selects which muscles to contract.

Manipulation Techniques

Different methods have been proposed to manipulate virtual agents: key-frame, script language, "performance animation" and task specification.

The key-frame technique. Key-frame requires a complete description of each frame of activity. The user places each object in the virtual world and has total control of their location and position. The main disadvantage of this method is that the total specification of the model requires immense amounts of data.

Script Language. Script language offers the possibility of performing complex animations (Kalra, Mangili, Magnenat-Thalmann, & Thalmann, 1991; Moravetz, 1989). Detailed lists of actions – in parallel or sequentially -- and their location and duration are specified. Examples of scripts include smile while saying “hello,” or start the action “walk” at time t, start action “wave hand” at time t+1, end action “wave hand” at time t+2. Script language provides a simple mechanism for scheduling actions and their

sequences.

Performance animation. "Performance animation" consists of recording the movement of an actor or an object through the use of sensors (DeGraf, 1990; Patterson, Litwinowicz, & Green, 1991; Litwinowicz, 1994; Guenter, et al, 1998). For example, sensors are placed on various points on the person being tracked. The movements of the points over time are used as input for a 3D synthetic model. The synthetic model moves by imitation.

This technique is mainly used in advertising and entertainment. Its main advantage is to produce complex animations quickly and cheaply. However each new animation requires new data. The synthetic agent has no knowledge simply reproducing the motions recorded.

Task specification. The task specification approach allows the user to give task-level instructions to an agent: "Go to the wooden door and open it." The program decomposes tasks into sub-goals (walk to the door, avoid any obstacle, find the type of door, grab the handle, open the door depending on its type (slide it or turn the knob and push the door)). Each sub-goal must be programmed using lower level functions: e.g. walking, grasping (Brooks, 1991; Zeltzer, 1991; Webber et al, 1995). The agent needs to evaluate and understand a situation (Chopra-Khullar & Badler, 1999) and must make decisions based on world knowledge and current goals.

Simulating Conversation between Agents

Communication in face-to-face interactions is expressed through a variety of channels, including the body, the voice, the face, and the eyes. When talking, humans move their hands (beats, batons, deitics) and heads (nods on accented items, gaze at the listener during back-channel) among other things. They accentuate words, and raise their eyebrows to punctuate a question mark or express affect. Speakers use facial expression, gaze and gesture not only to reinforce their talk but also to convey their emotion and to evaluate their partner's reaction. Moreover, these non-verbal signals are synchronized with the dialogue and with the agent's activity (gaze follows hand movement while performing a task).

To have a believable animation a synthetic agent must deploy each of these behaviors in a way that is appropriate and well-timed.

Face-to-face conversation between synthetic agents. The goal of many simulations (Cassell et al. 1994) is to simulate interaction in which one agent helps the other to achieve a goal. Each agent is implemented as semi-autonomous keeping its own representation of the state of the world and the conversation, and whose behavior is determined by these representations. The appropriate intonation, gesture, gaze and facial expressions are computed based on the semantic content and the dialogue generated by a discourse planner.

In this model the two agents do not sense each other's behaviors. This is a significant limitation because responsive interactions require dynamic adjustments to each agent's behaviors. Without sensing the partner's behavior, no adjustment by the agent to ongoing actions by the partner is possible. Instead the complexities of this version are found in the coordination within an agent's behavioral systems rather than between agents.

Face-to-face conversation between a synthetic agent and a user. Takeuchi and Nagao (1993; Nagao & Takeuchi, 1994) move a step closer to realistic responsive interactions. They employ a categorization of facial expressions that depends on communicative meaning. Chovil (1991) postulates that facial expressions are not only a signal of the emotional state of the sender but also a social communication whose conveyed meanings have to be interpreted in the context in which the expressions are emitted. She found that facial displays occurring during speech are linked to current semantic content.

Based on these insights, the authors consider twenty-six facial displays stored in a library. When a response is computed in the speech dialogue module, a corresponding facial display is generated simultaneously. A signal is sent to the animation module, which deforms the facial model to show the requested facial displays. In Takeuchi model, the facial actions of agent B depend on the semantic

content presented by agent A. Although simplistic, there is a rudimentary form of responsiveness with agent A's actions dependent on those of B.

Most recent conceptual advances include the development of the embodied agent -- that is one encompassing conversational skills and able to exhibit nonverbal communicative behaviors (Andre et al, 2000; Badler et al, 2000; Cassell et al, 2000; Rickel and Johnson, 2000; Lester et al, 2000; Poggi and Pelachaud, 2000, Poggi et al, 2000)). The goal of this work is to develop an agent capable of understanding the user's verbal and nonverbal behaviors, as well as being able to generate human-like communicative behaviors.

Ymir (Thórisson, 1997) is an architecture to simulate face-to-face conversation between the agent, Gandalf, and a user. The system takes as sensory input hand gesture, eye direction, intonation, and body position of the user. Gandalf's behavior is computed automatically in real time. He can exhibit context-sensitive facial expressions, eye movement, and pointing gestures as well as generate turn-taking signals. Nevertheless, Gandalf has limited capacity to analyze the discourse at a semantic level and therefore to generate semantically driven nonverbal signals.

Rea, the real estate agent, is capable of multimodal conversation: she can understand and answer in real time (Cassell, et al, 1999). She moves her arms to indicate and to take turns. She uses gaze, head movements, and facial expressions for functions such as turn taking, emphasis, and greetings as well as for back channel to give feedback to the user speaking to her. Poggi and Pelachaud (2000) developed a system of an animated face that can produce the appropriate facial expression according to the performative of the communicative act being performed, while taking into account information on the specific interlocutor and the specific physical-social situation at hand.

## Conclusions and Future Directions

We have reported different techniques to simulate complex animations and behaviors during conversation. They offer tools to analyze, manipulate and integrate systems so that models of

communication between agents can be realistic. But to take full advantage of these techniques in simulating human interaction requires clear ideas about how humans interact in general and in the specific context of cooperative exchanges.

Current simulations of social interaction have a variety of shortcomings. The interface between the synthetic agent interacting with a human requires a better sensing and recognition system. Current systems limit the role of humans to simple spoken utterances with some head and hand motions, as well as a few facial expressions. Moreover, while dialoging with a synthetic agent, most of the time no interruption by the user is allowed. (however see Cassell et al, 1999). Also the set of utterances used by the system is small.

We believe that successful models require not only production and recognition systems, not only coordination among gestural, vocal, and semantic subsystems, but also models that incorporate responsive agents. Responsiveness implies the ability to adjust to the dynamically changing behavior of the partner in ways that mimic at least approximately the alterations that humans would make to one another in similar, usually cooperative contexts.

**Responsiveness in Human Social Interaction**

A comprehensive model of human social interaction would include both semantic and emotional components. In the data presented here only emotional components will be considered. Human emotion is carried in a variety of ways in social interaction but the nonverbal channel including face, voice, and body is the primary vehicle of emotional communication (Cappella, 1991). Social attachment and affective reaction are conveyed and understood in the patterns of emotional signaling through the voice (Scherer, 1986) and face (Ekman, 1971) as well as body position (Hatfield, Cacioppo, & Rapson, 1994) and less observable physiological indicators (Ekman, Levenson, & Friesen, 1983).

In this section our attention will be focused on the ways that nonverbal signs of affect are expressed and responded to in ordinary social interaction. By understanding the patterns of exchange

and response between humans in cooperative interactions, we hope to be able to infer some specific and general rules for virtual interaction.

Much of the information that researchers have gathered about human interaction is based on static data or, at best, scenarioes in which two exchanges are monitored. The data to be reported here comes from interactions that take place over 20 to 30 minute periods. The behaviors enacted in those periods are audio and video recorded for later coding.

The archive of interactions we have consists of about 100 interactions. They include same sex and opposite sex pairs, dyads with longer histories (greater than six months as friends) and strangers, partners with similar and different attitudes, and expressive and reticent pairs (see Cappella & Palmer, 1990 or more details on the design and procedures for data collection). This group of persons offers maximum variance of behavioral response in part due to their expressive differences. Their interactions were informal and not directed by the researcher in any way. The interactions scrutinized in this paper come from a set of 19 interactions of 15 minute duration.

A number of behaviors were coded for later analysis. These include vocalic behaviors, eye gaze, smiles and laughter, head nods, back channels, posture, illustrator gestures, and adaptor gestures. Vocal behaviors allow us to obtain information about conversational tempos that are known to be related to arousal and excitation. Overlapping speech patterns can be read as impolite as people are seen to usurp conversational resources. Positive affect is carried by in part by facial smiles and laughter. Head nods provide feedback while listening as well as emphasis during speech. Gaze can be a regulator of interaction, a method of monitoring threat, or a sign of attention, and positive regard. Gestures can function as signs of anxiety and spillover of energy and as a means of carrying information that is redundant with or supplementary to speech. Back channels are signals listeners offer speakers that they are being attentive while not necessarily trying to wrest the floor away. Postural states may be signals of involvement or of detachment.

Behaviors are carefully and reliably assessed using trained coders and computerized data acquisition techniques.[1] Codes are "on and off" values at each 0.1 second yielding long time series for each behavior and each person. The series are synchronous with a common time base. These series give a temporally precise picture of the behaviors enacted by partners during ordinary social interaction. Since some of these behaviors carry information about affect, they provide the basis for describing emotional responsiveness.

Analytic Strategy

The long term goal of our research is to model the sequential structures of human interaction, specifically the behaviors indicative of emotional reaction. Our approach identifies states of the individual and the interaction. Writing rules that describe changes in these states over time and that correspond to the empirical realities is the essence of the enterprise of modeling. Consider the case of smiling and the rules that might govern its enactment.[2]

To describe interaction, two types of rules need to be understood. One set concerns sequence or when to change a state. For example, do people break mutual gaze by both looking away at the same time or does one look away first? The other concerns distributional rules or how long to remain in a state before leaving. For example, how frequent is a gaze of more than 6 seconds? Is this a common or uncommon occurrence? Because these rules are probabilistic, the range of observed probabilities can provide guidance to modelers about what humans find acceptable and unacceptable changes in behavior during interaction.

A second issue concerns the source of probabilities for rules of sequence and distribution. Can we study the behavior of individuals to see how and when they change or must we focus on the behavior of pairs of persons within interaction? Are interactions homogeneous regarding distributional and sequential rules or do the rules change from one section to the next? This is sometimes called context sensitivity. Are interactional rules context sensitive or not? We will take up each of these questions in

turn.

Rules from the Behavior of Individuals

In Table 1, probabilities of individual change in four behaviors are presented. The behavior is assumed to be either "on" or "off". The matrix is the probability of moving from a prior to a subsequent state. These probabilities are derived from treating each person in the interaction as if he or she did not have a partner. The cell of each matrix contains an average probability and a high and low value. The number of observations is more than 300,000.

Two things are immediately apparent. First, some behaviors are much more frequent than others. Body gestures occur at roughly 45% of the time while smiles and illustrator gestures are "off" the vast majority of the time. Gaze directed at the partner is on at the rate of 80% on average. Second, there is considerable variability across persons. The high and low values can differ by huge amounts, at times spanning the full range of probabilities.

What is not so obvious from these data are their implications for responsiveness. Can individual transition probabilities be used to create sequences for pairs of people in interaction? The answer is no on two grounds. The variability in individual response implies that the average values will not provide good fit for any particular dyad. Also when two people are paired in interaction, there is good evidence that they adjust their behaviors to those of the partner, for example, in cooperative interactions smiling together and converging in their interactive tempos (Burgoon, Stern, & Dillman, 1995; Cappella, 1981, 1991). This implies that we cannot predict well A's interaction with C based on A's interaction with B and C's interaction with D (Cappella, 1980).

The first rule of interaction, then, is the synthesis rule. The behavior of persons is insufficient for synthesizing the behaviors of dyads. Studying the behaviors of individuals can never produce realistic descriptions of dyads. Put a bit more technically, the probabilities that describe a dyad when derived from the probabilities that describe persons will yield unrealistic models of interaction (virtual or

otherwise).[3]

Table 1 about here

Predicting Sequential Rules from Dyads

In order to study the sequences of behavior in dyads, we first need to create state definitions for pairs of people in interaction. If these descriptions are to avoid the synthesis problem, then they must be sensitive to the behavior of the partner and not just the behavior of the person. The usual means for doing so is is to define states for the pair of persons as follows:

State Definitions For Any 2-Person, On-Off Behavior (Example for Smiles)

| A's Behavior | B's Behavior | Dyad's Behavior |
|---|---|---|
| Smile is off (=0) | Smile is off (=0) | NEITHER Smiling (00) |
| Smile is on (=1) | Smile is off (=0) | A ONLY Smiling (10) |
| Smile is off (=0) | Smile is on (=1) | B ONLY Smiling (01) |
| Smile is on (=1) | Smile is on (=1) | BOTH Smiling (11) |

Using these state definitions, we can follow the sequences among the various dyadic states. These are represented by transition matrices but now the transition matrices describe movements by pairs of people over time rather than individuals changing. A matrix representing 19 different dyads aggregated together is presented in Table 2.

What do transition matrices tell us? First, the diagonal elements (upper left to lower right) indicate the probability that the dyad continues in the state that it is already in. The off-diagonal elements tell us about changes from one condition to the next – for example, from only person A smiling

to both A and B smiling together. In effect, the diagonal elements give information about stability of a state while the off-diagonals give information about change.

Table 2 about here

Let us work with the case of smiling and laughter because this is a crucial variable in some later studies we will be discussing. Smiles and laughter are mostly off for the dyad. When the dyad changes state the paths it does not take include

00 → 11 (Neither → Both)

10 → 01 (A only → B only)

01 → 10 (B only → A only)

11 → 00 (Both → Neither)

That is, the cross-diagonals (lower left to upper right) are zero. People do not change from one person smiling alone to the other smiling alone or from both smiling to neither smiling or neither to both smiling. In human terms, they negotiate.

Instead to get from one mutual state to another or to get from one person smiling alone to another smiling alone the following paths are used.

00 → 01 → 11

OR

10 → 11

10 → 00 → 01

OR

11 → 01

01 → 00 → 10

OR

11 → 10

11 → 01 → 00

OR

10 → 00

When neither is smiling, an overture by one is required before acceptance by the other is possible. When both are smiling termination by one is required before termination by both. Most interestingly, smiling by one can only become smiling by the other through moments of mutuality. What does not happen is alternation of smiling alone or a sequence when smiling together follows neither smiling or neither smiling follows smiling together. This kind of dyadic behavior appears to be forbidden in human interaction and, therefore, should be forbidden in virtual interactions as well.[4]

Two conclusions obtain. First, mutuality is a crucial state for how the dyad changes its conditions of smiling. Second, to have mutuality requires a person knowing his or her own state as well as that of the partner. A realistic model of smiling in interaction cannot be built from studying the behavior of individuals or through simple sequences of expression and recognition guided by individual rules.

When other behaviors are examined such as gaze and gesture, patterns similar to those observed with smiles are found. Adaptor gestures show the greatest variability with the diagonal probabilities varying from very low to very high. Behaviors that are mostly on (e.g. gaze) and mostly off (e.g.

gestures) have smaller ranges of variation.

In general, the dyadic matrices exhibit more empirical constraint than the individual matrices do. The off-diagonal probabilities carry information about changes in the state of the pair of persons. In all cases, the cross-diagonal elements are zero or nearly zero. This constraint implies that when the dyad changes state it does so along a particular path and avoids other paths completely. The paths people choose are through moments of sharing the same state. This simultaneity is a kind of mutual responsiveness that is not required in principle but is required by the social nature of human beings.

Context Effects

In the study of grammars, one distinguishing feature of types of grammar is whether they are context sensitive or not. Are there features of the surrounding linguistic context that determine the application of one rather than another rule? Context sensitivity may also apply to the study of the grammar of emotional exchange.

An important context in all interactions is the exchange of speaker and hearer roles, also called turn-taking (Duncan & Fiske, 1977). Speakers and hearers are different behaviorally in many ways. Speakers are generally under greater cognitive load than listeners are (Cappella, 1980). They look at listeners less and, of course, gesture more (Cappella, 1985). The kinds of head nods used are very different tending to be more related to packets of stressed speech than the deliberate nods of listeners (Duncan & Fiske, 1977). Too, holding the floor is controlling an important conversational resource that must be shared or, if not, wrestled away from the partner in order to again access.

The listener-hearer role may be one important context within which other social and emotional exchanges occur. To determine whether sequential rules are context sensitive, we first need to define states and sequences of states for two person speaker-hearer exchanges and, then, embed social-emotional exchange rules into these contexts.

Table 3 about here

States for turn taking are presented in Table 3 and are based on the definitions of Jaffe and Feldstein (1970). The definitions depend on two important features. First is that having the floor is the same as being the only speaker. Second, a person has the floor from the person's first unilateral vocalization to the first unilateral vocalization by the partner. In Table 3 there are 6 rather than 4 dyadic states of previous representations. This is because "holding the floor" is ambiguous when both are silent or both are talking. The ambiguity is resolved by giving the person who has most recently had the floor responsibility for the floor in subsequent moments of mutual silence or mutual talk.

With six speaker-hearer states, a first order transition matrix will have 36 (=6x6) cells. But some of these cells have structural zeros because certain sequences are forbidden by definition. For example, the dyad cannot change from both talking and person A holding the floor to both talking and person B holding the floor. In the 6x6 transition matrix, there are 12 such constraints (also called structural zeros).

To test for context sensitivity, the sequential matrices for emotion and social behavior must be embedded within the speaker-hear transition matrix producing a rather daunting 24x24 matrix with 12x16 (=192) structural zeroes. The general matrix is very complex and is only presented in the appendix. The complexity suggests that even simple codes for behavior (such as on and off) can quickly produce very involved representations just by requiring dyadic rather than individual representations and context sensitivity rather than context independence.

The complexity of context-sensitive affective exchanges can be reduced by noticing that certain transitions can be grouped conceptually. We divided the sub-components of this transition matrix into four speaker-listener contexts summarized in Table 4. They include the most common types of speaker-hearer exchanges: ordinary speaker exchanges; ordinary continuations of the speaker role; contests for the speaker role won by the original speaker; and awkward moments where it is not clear who will get the floor next.

Table 4 about here

Context sensitivity asks: Are the sequential rules for smiling, gesturing, and gazing the same or different across the contexts of speaker-hearer interaction? The summary matrix for smiles within the four speaker-listener contexts is presented in appendix A. First the composite matrix is listed followed by smile sequences during turn switches, simultaneous turns, within turns, and awkward turns.

The large sample sizes insure that the smile sequences are reliably different form the composite for the different contexts. The match between the composite and the smile sequences for the "within turn" context is very close mostly because 88% of the observations for the composite come from moments in the interaction when a person is continuing to hold the floor. The other 3 smile sequence matrices differ from the composite by amounts which can be appreciable.

Specifically, the row totals for smiling are higher during turn switches, awkward, or simultaneous turns then during within-turn interaction. Although the data are not presented here, there is more mutual gaze during turn switches, simultaneous turns, and awkward turns than during within-turn segments. In effect, smiling and mutual gaze tend to pile up during those moments in interaction when speaker-listener roles are being exchanged, the roles are being contested, or when awkward moments such as an attempted interruption followed by mutual silence. By contrast, when speakers are engaged in serial monologue, mutual smiling is lowered. To put too simple a point on these data: social and emotional rules of interaction depend on turn-taking context.

Sometimes the differences described in the above sections appear to be rather small in magnitude. However, both participants in and observers of interactions use these differences in responsiveness in the judgements they render about interlocutors.

Being Micro-responsive Matters.

One could respond to our findings so far as "much ado about nothing." Small changes like these could not matter much to ordinary interactants. We undertook a series of studies to test whether the micro processes of interaction matter.

From the 100 or so dyadic interactions in our archive, eight were selected. Four of these met criteria for highly responsive interactions and four were low in responsiveness. Responsiveness was defined using time series methods with equations similar to equations 1 and 2 presented earlier. From these eight, two one minute segments from each were chosen (see Cappella, 1997 for further details).

Three studies were conducted. The first simply showed the 16 one-minute segments in a fixed order. People evaluated each immediately after seeing the segments. Four questions were asked, each assessing some component of responsiveness. In a second study, facial cues were removed by superposing a mosaic on the faces. Motion was still visible but specific features were not. In a third study, both facial and vocal cues were eliminated. Vocal cues were completely eliminated in study 3 while in study 1 words could not be understood although vocal tempo and variation could be.

Students in study 1 denied that they could make reliable judgments of responsiveness. There were incorrect in their denials because judgments were reliable within person, within study, and across studies. People were sensitive to responsive interactions being able to distinguish responsive from non-responsive interaction in all three studies. Observers judged partners to be responsive when they smiled in synchrony with one another and when their gaze and gesture were complementary. One way of describing this is that partners were judged synchronous when they were emotionally responsive and polite. Interactants liked one another more when their smiles were mutual ones. Judged responsiveness too accounted for people's attraction to one another.

The implications of these results are, we think, very important for building virtual interactions. If people are going to judge virtual social interactions as real, then simulations must be sensitive to micro-momentary responsiveness and unresponsiveness between partners. People are sensitive to responsive partners whether they are participating in the interaction or just observing it. They may not be able to say what it is about an interaction that makes it feel right or wrong but they do perceive unresponsive partners in less favorable terms.

## Conclusion and Implications

Among researchers in the AI community, there has been a sharp upsurge of interest in creating synthetic agents with at least some capacity for interaction with human agents. Many researchers (e.g. Picard, 1997) have argued that computerized tools need to be "emotionalized" in part because people feel comfortable treating computers and other media in social terms and in part because emotion is as important a component of the learning process as rationality is. Making computers, or their virtual agents, more user-friendly involves adaptation in both rational and emotional ways.

The task of creating emotionally responsive synthetic agents is enormously complex. Multiple systems must be coordinated within a given synthetic agent just to make the agent's actions appear roughly normative. These subsystems include the semantic, vocal, gestural, facial, visual, and so on. However, to fabricate a synthetic agent with the capacity for interaction with a human or another synthetic agent requires responsiveness between agents. And responsiveness between agents is more than a sequence of interleaved expressions, no matter how realistic those expressions might be. Realistic virtual interactions require agents responsive to one another's behavior just as human interaction, if it is to be human, requires responsiveness between partners.

Many of the tools employed in AI modeling efforts make assumptions that simplify the processing load by avoiding the inclusion of recognition systems or building in pre-established goals and plans. These simplifications are understandable at the earlier stages of modeling. However, simulations that produce realistic virtual interactions will need to include agents with the ability to sense their own state as well as that of the partner and the capacity to dynamically alter their behavior in response to that of the partner and to the surrounding context.

Our data from the human sphere made very clear that interactions cannot be modeled by studying the behavior of individuals disaggregated from their partners. Rather, partners must be studied together. You cannot build models of dyads from the behavior of individuals. The reason is simply that people

adjust to their partners' behaviors – that is they are responsive. There is an aggregation problem in moving from persons to dyads.

People are also sensitive to the context of their actions. For example, smiling (and gazing) were more frequent when partners were switching speaker and listener roles or contesting those roles than when carrying out a lengthy monologue. Virtual agents will need the capacity to know what context their actions are in so that minor modifications in affective cues such as smiles can be made.

The perceived realism of an interaction depends in part on these micro-adjustments. Humans who participate in or observe interactions that involve less responsive others sense it and evaluate the interaction less favorably. Although current synthetic agents may behave in ways that are too crude to worry about micro-adjustments in smiles, gaze, gestures, and head nods, eventually they will need to. The models employed as the tools for simulation will require assumptions that allow for responsive, context-sensitive agents.

The study of interpersonal relations is about to face a new set of entities for its empirical and theoretical scrutiny. These entities will be the robots, virtual and synthetic agents that will interact with one another and with human agents. Whether the tools used in the study of personal and social relationships will be useful in this new domain of relationships is unclear. What is clear is that scholars of interpersonal relations have the opportunity not only to study but to participate in the creation of the objects of study.

Bibliography

Andre, E., Rist, T., van Mulken, S., Klesen, M., & Baldes, S. The Automated design of Believable Dialogues for Animated Presentation Teams. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), Embodied Conversational Agents. Cambridge, MA: MIT Press

Badler, N., Bindiganavale, R., Allbeck, J., Schuler, W., Zhao, L., & Palmer, M. Parameterized Action Representation for Virtual Human Agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), Embodied Conversational Agents. Cambridge, MA: MIT Press.

Brooks, R.A. (1991). A robot that walks: Emergent behaviors from a carefully evolved network. In N. I. Badler, B.A. Barsky, & D. Zeltzer (Eds.), Making them move: Mechanics, control, and animation of articulated figures,(pp. 99-108). San Mateo, CA: Morgan-Kaufmann.

Burgoon, J. K., Stern, L.A., & Dillman, L. (1995). Interpersonal adaptation: Dyadic interaction patterns. New York: Cambridge University Press.

Cappella, J.N. (1980). Talk and silence sequences in informal social conversations II. Human Communication Research, 6, 130-145.

Cappella, J.N. (1981). Mutual influence in expressive behavior: Adult and infant-adult dyadic interaction. Psychological Bulletin, 89, 101-132.

Cappella, J.N. (1985). Production principles for turn-taking rules in social interaction: Socially anxious vs. socially secure persons. Journal of Language and Social Psychology, 4, 193-212.

Cappella, J. N. (1991). The biological origins of automated patterns of human interaction. Communication Theory, 1, 4-35.

Cappella, J. N. (1993). The facial feedback hypothesis in human interaction: Review and speculations. Journal of Language and Social Psychology, 12, 13-29.

Cappella, J.N. (1994). The management of conversational interaction in adults and infants. In M.L. Knapp & G.R. Miller (Eds.), The handbook of interpersonal communication (2nd ed., pp. 380-419).

Thousand Oaks, CA: Sage.

Cappella, J. N. (1997). Behavioral and judged coordination in adult informal social interactions: Vocal and kinesic indicators. Journal of Personality and Social Psychology, 72, 119-131.

Cappella, J.N. & Palmer, M.T. (1990). Attitude similarity, relational history, and attraction: The mediating effects of kinesic and vocal behaviors. Communication Monographs, 57, 161-183.

Cassell, J.B., Campbell, J., Vilhjalmsson, L., & Yan, H. Human Conversation as a system framework: Designing embodied conversational agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), Embodied Conversational Agents. Cambridge, MA: MIT Press

Cassell, J., Bickmore, J., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., & Yan, H. (1999). Embodiment in conversational interfaces: Rea. In CHI '99 Conference Proceedings, 520-527, Pittsburgh, Pennsylvania.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn B., Becket T., Douville B., Prevost S., & Stone M. (1994). Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. Computer Graphics Annual Conference Series, 413-420.

Chopra, S.K. & Badler, N. (1999). Where to look? Automating attending behaviors of virtual human characters. In Proceedings of Autonomous Agents '99, Seattle, Washington.

Chovil, N. (1991). Social determinants of facial displays. Journal of Nonverbal Behavior, 15, 141-154.

Davis, D. & Holtgraves, T. (1984). Perceptions of unresponsive others: Attributions, attraction, understandability, and memory for utterances. Journal of Experimental Social Psychology, 20, 383-408.

Davis, D. & Martin, H. J. (1978). When pleasure begets pleasure: Recipient responsiveness as a determinant of physical pleasuring between heterosexual dating couples and strangers. Journal of Personality and Social Psychology, 36, 767-777.

Davis, D. & Perkowitz, W.T. (1979). Consequences of responsiveness in dyadic interaction: Effects of probability of response and proportion of content-related responses on interpersonal attraction. Journal of Personality and Social Psychology, 37, 534-550.

DeGraf, B. (1990). "Performace" facial animation. In State of the Art in Facial Animation (vol. 26), 10-14. ACM Siggraph'90.

Duncan, S.D., Jr. & Fiske (1977). Face-to-face interaction: Research , methods, and theory. Hillsdale, NJ: Lawrence Erlbaum.

Ekman, P. (1971). Universal and cultural differences in facial expressions of emotion. In J. Cole (Ed.), Nebraska symposium on motivation (pp. 207-283). Lincoln: University of Nebraska Press.

Ekman, P., Levinson, R.W., & Freisen, W.V. (1983). Autonomic nervous system activity distinguishes among emotions. Science, 221, 1208-1210.

Gardner, H. (1983). Frames of mind. New York: Basic Books.

Gardner, H. (1995). Multiple intelligences: The theory in practice. New York: Basic Books.

Goleman, D. (1995). Emotional intelligence. New York: Bantam Books.

Guenter, B., Grimm, C., Malvar, H. & Wood, D. (1998). Making faces. Computer Graphics Proceedings, Annual Conference Series, ACM.

Jaffe, J. & Feldsetin, S. (1970). Rhythms of dialogue. New York: Academic Press.

Kalra, P., Mangili, A., Magnenat-Thalmann, N., & Thalmann, D. (1991). SMILE: A multilayered facial animation system. In T. L. Kunii (Ed.), Modeling in Computer Graphics. Springer-Verlag.

Ko, H. (1994). Kinematic and Dynamic Techniques for Analyzing, Predicting, and Animating Human Locomotion. Ph. D. Dissertation, University of Pennsylvania, Philadelphia, PA.

Lee, Y., Terzopoulos, D., & Waters, K. (1995). Computer Graphics Annual Conference Series, 1995.

Lester, J., Towns, S., Callaway, C., Voerman, J., & FitzGerald, P. Deictic and Emotive Communication in Animated Pedagogical Agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), Embodied Conversational Agents. Cambridge, MA: MIT Press.

Litwinowicz, P.C. (1994). Animating images with drawings. Computer Graphics Annual Conferences Series, 413-420.

Magnenat-Thalmann, N. & Thalmann, D. (1987). The direction of synthetic actors in the film: Rendez-vous a Montreal. IEEE Computer Graphics and Applications, December, 1987, 9-19.

Miller, M. J. (1999). Computers will be more human. PC Magazine, June 22, 1999.

Moravetz, C. (1989) A high level approach to animating secondary human movement. Master's thesis, School of Computing Science, Simon Fraser University.

Nagao, K. N. & Takeuchi, A. (1994). Speech dialogue with facial displays: Multimodel human-computer conversation. In ACL '94, 102-109.

Negroponte, N. (1996). Affective Computing. Wired, April, 1996.

Patterson, E.C., Litwinowicz, P.C., & Greene, N. (1991). Facial animation by spatial mapping. In N. Magnenat-Thalmann & D. Thalmann, editors, Computer Animation '91, 45-58. Springer-Verlag.

PC Magazine (1999). Emotional computing. July 1999.

Phillips,C.B. & Badler, N.I. (1991). Interactive behaviors for articulated figures. Computer Graphics, 25, 359-362.

Picard, R. (1997). Affective computing. Cambridge, MA: MIT Press.

Poggi, I., & Pelachaud C. (2000). Performative Facial Expressions in Animated Faces. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), Embodied Conversational Agents. Cambridge, MA: MIT Press.

Poggi, I., Pelachaud, C., & de Rosis, F. (2000). Eye communication in a conversational 3d synthetic agent. Special Issue on Behavior Planning for Life-Like Characters and Avatars of AI

Communications, 2000.

Reeves, B. & Nass, C. (1996). The media equation. New York: Cambridge University Press.

Rickel, J. & Johnson, W.L. Task-Oriented Collaboration with Embodied Agents in Virtual Worlds. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), Embodied Conversational Agents. Cambridge, MA: MIT Press

Rijpkema, H. & Girard, M. (1991). Computer animation of hands and grasping. Computer Graphics, 25, 339-348.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. Psychological Bulletin, 99, 143-165.

Takeuchi, A. & Nagao, K. (1993). Communicative facial displays as a new conversational modality. In ACM/IFIP INTERCHI'93, Amsterdam, 1993.

Thòrisson, K.R. (1997). Layered modular action control for communicative humanoids. In Computer Animation'97, Geneva: IEEE Computer Society Press.

Webber, B., Badler, N., Di Eugenio, B., Geib C., Levison, L., & Moore, M. (1995). Instructions, intentions and expectations. Artificial Intelligence Journal, 73, 253-269.

Zeltzer, D. (1991). Task-level graphical simulation: Abstraction, representation, and control. In , N.I. Badler, B.A. Barsky & D. Zeltzer (Eds.), Making them move: Mechanics, control, and animation of articulated figures (pp. 3-33). San Mateo, CA: Morgan-Kaufmann.

Zhao, J. & Badler, N. I. (1994). Inverse kinematics positioning under nonlinear programming for highly articulated figures. ACM Transactions on Graphics, 13, 313-336.

Table 1. Transition probability matrix for four behaviors at individual level: Composite, low, and high values.

| ADAPTORS | OFF | ON | TOTAL |
|---|---|---|---|
| OFF: Average | .5362 | .0047 | .5410 |
| (Low-High) | (.0980-.9516) | (.0014-.0140) | |
| ON: Average | .0047 | .4543 | .4590 |
| (Low-High) | (.0013-.0140) | (.0400-.8978) | |
| GAZE | | | |
| OFF: Average | .1854 | .0136 | .1990 |
| (Low-High) | (.0100-.3538) | (.0027-.0224) | |
| ON: Average | .0135 | .7873 | .8008 |
| (Low-High) | (.0026-.0226) | (.6419-.9630) | |
| ILLUSTRATORS | | | |
| OFF: Average | .9128 | .0044 | .9172 |
| (Low-High) | (.6207-.9827) | (.0002-.0128) | |
| ON: Average | .0044 | .0784 | .0828 |
| (Low-High) | (.0002-.0128) | (.0009-.2198) | |
| SMILES | | | |
| OFF: Average | .9274 | .0034 | .9308 |
| (Low-High) | (.8566-.9841) | (.0011-.0077) | |
| ON: Average | .0034 | .0658 | .0692 |
| (Low-High) | (.0011-.0077) | (.0133-.1451) | |

Note. The first entry in each cell is the average probability across 38 people; second is the lowest and third the highest probability.

Table 2. Transition probability for dyadic state: Average, high and low values for SMILES & LAUGHTER (N=170586).

| | Neither on | A on only | B on only | Both on | Row Total |
|---|---|---|---|---|---|
| Neither on | .8830 | .0029 | .0018 | .0001 | .8878 |
| | .7341-.9468 | .0009-.0073 | .0002-.0046 | 0-.0004 | |
| A on only | .0029 | .0544 | .0000 | .0011 | .0583 |
| | .0009-.0072 | .0071-.1228 | 0-.0001 | .0003-.0020 | |
| B on only | .0019 | .0000 | .0253 | .0007 | .0279 |
| | .0003-.0050 | 0-.0001 | .0027-.0548 | .0001-.0014 | |
| Both on | .0000 | .0011 | .0008 | .0241 | .0260 |
| | 0-.0004 | .0004-.0021 | .0002-.0022 | .0071-.0797 | |

Table 3. Defining speaker-listener-states according to the rules of Jaffe and Feldstein (1970).

| PERSON A SPEAKING | PERSON B SPEAKING | FLOOR? | STATE DESCRIPTION | STATE CODE |
|---|---|---|---|---|
| NO | NO | A | BOTH SILENT A FLOOR | 00A |
| YES | NO | A | A ONLY | 10A |
| YES | YES | A | BOTH TALK A FLOOR | 11A |
| NO | NO | B | BOTH SILENT B FLOOR | 00B |
| NO | YES | B | B ONLY | 01B |
| YES | YES | B | BOTH TALK B FLOOR | 11B |

Table 4. Four speaker – listener contexts that may alter emotional interaction patterns.

| CONTEXT | ELEMENTS OF CONTEXT | STATE CHANGES |
|---|---|---|
| SWITCHING SPEAKER & LISTENER ROLES | SMOOTHLY W/ SWITCHING PAUSE | 10A → 01B |
| | SMOOTHLY W/O SWITCHING PAUSE | 00A → 01B |
| | INTERRUPTIVE W/O SWITCHING PAUSE | 11A → 01B |
| SIMULTANEOUS CONTESTS FOR SPEAKER ROLE | CONTESTING | 11A → 11A |
| | END CONTESTING | 11A → 10A |
| | BEGIN CONTESTING | 10A → 11A |
| WITHIN SPEAKER ROLE | NORMAL CONTINUATION | |
| | WITH SPEECH | 10A → 10A |
| | WITHOUT SPEECH | 00A -→ 00A |
| | END HESITATION | 00A → 10A |
| | BEGIN HESITATION | 10A 0→0A |
| AWKWARD MOMENTS: | WHOSE TURN? | 11A → 00A<br>00A → 11A |

Appendix A

Two tables follow. Table A.1 is a transition matrix for smiles within context. The contexts are determined by speaker-hearer roles in conversation and transitions between those roles. Embedded within each role and role transition are dyadic sequences for smiling and laughter. Table A.2 is a set of 5 matrices. The first is the composite matrix for dyadic smile sequences, identical to that presented in earlier tables. The next four are the matrices for the same behavior and same sequences but in the context of switching between speaker and hearer roles; simultaneous speaking; within-turn speaking; and awkward turns.

Table A 1. Transition matrix necessary to detect context sensitivity of behavioral sequences: Example of smile.

| Floor | Smile | 00 A | | | | 10 A | | | | 11 A | | | | 00B | | | | 01 B | | | | 11B | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 00 | 10 | 01 | 11 | 00 | 10 | 01 | 11 | 00 | 10 | 01 | 11 | 00 | 10 | 01 | 11 | 00 | 10 | 01 | 11 | 00 | 10 | 01 | 11 |
| **00A** | **00** | | wt | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **00A** | **10** | | wt | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **00A** | **01** | | wt | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **00A** | **11** | | wt | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **10A** | **00** | | | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **10A** | **10** | | | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **10A** | **01** | | | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **10A** | **11** | | | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **11A** | **00** | | | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **11A** | **10** | | | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **11A** | **01** | | | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **11A** | **11** | | | | | | | | | | | | | x | x | x | x | | | | | x | x | x | x |
| **00B** | **00** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **00B** | **10** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **00B** | **01** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **00B** | **11** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **10B** | **00** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **10B** | **10** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **10B** | **01** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **10B** | **11** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **11B** | **00** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **11B** | **10** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **11B** | **01** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |
| **11B** | **11** | x | x | x | x | | | | | X | x | X | x | | | | | | | | | | | | |

Figure: A and B refer to agent1 and to agent2; floor: 0 pause, 1:talk, floor to agent A or B; smile: 1: smile or laughter, 0: no smile or laughter; wt:

within-turn. “x” implies forbidden transition.

Table A2. Transition matrices for smiles: Composite and by context of occurrence

SMILE

COMPOSITE (N=170586)

| | NEITHER | A ONLY | B ONLY | BOTH | Total |
|---|---|---|---|---|---|
| Neither | 0.882951 | 0.002861 | 0.001799 | 0.000141 | 0.887752 |
| A Only | 0.002854 | 0.054371 | 0.000018 | 0.001070 | 0.058304 |
| B Only | 0.001883 | 0.000006 | 0.025320 | 0.000691 | 0.027900 |
| BOTH | 0.000059 | 0.001079 | 0.000761 | 0.024130 | 0.026029 |

Turn Switches (N=4590)

| | NEITHER | A ONLY | B ONLY | BOTH | Total |
|---|---|---|---|---|---|
| Neither | 0.849455 | 0.006100 | 0.002614 | 0.000218 | 0.858387 |
| A Only | 0.003922 | 0.067756 | 0.000000 | 0.000871 | 0.072549 |
| B Only | 0.002832 | 0.000000 | 0.033987 | 0.001089 | 0.037908 |
| BOTH | 0.000000 | 0.001089 | 0.000871 | 0.029194 | 0.031154 |

Simultaneous Turn (N = 15562)

| | NEITHER | A ONLY | B ONLY | BOTH | Total |
|---|---|---|---|---|---|
| Neither | 0.839352 | 0.006297 | 0.003920 | 0.000386 | 0.849955 |
| A Only | 0.003084 | 0.069207 | 0.000000 | 0.002121 | 0.074412 |
| B Only | 0.002185 | 0.000064 | 0.039070 | 0.001349 | 0.042668 |
| BOTH | 0.000000 | 0.001542 | 0.001285 | 0.030138 | 0.032965 |

Within Turn (N=150252)

| | NEITHER | A ONLY | B ONLY | BOTH | Total |
|---|---|---|---|---|---|
| Neither | 0.888581 | 0.002396 | 0.001551 | 0.000113 | 0.892641 |
| A Only | 0.002802 | 0.052385 | 0.000020 | 0.000965 | 0.056172 |
| B Only | 0.001824 | 0.000000 | 0.023614 | 0.000612 | 0.026050 |
| BOTH | 0.000067 | 0.001032 | 0.000705 | 0.023334 | 0.025138 |

Awkward Turn (N=182)

| | NEITHER | A ONLY | B ONLY | BOTH | Total |
|---|---|---|---|---|---|
| Neither | 0.807692 | 0.010989 | 0.005495 | 0.000000 | 0.824176 |
| A Only | 0.000000 | 0.087912 | 0.000000 | 0.000000 | 0.087912 |
| B Only | 0.000000 | 0.000000 | 0.043956 | 0.000000 | 0.043956 |
| BOTH | 0.000000 | 0.000000 | 0.000000 | 0.043956 | 0.043956 |

End Notes

[1] Reliabilities are reported in Cappella and Palmer (1990) or are available upon request from the author.

[2] In this chapter space limitation require that we focus on only one behavior. We have selected smiles. Interested parties may contact the author for similar analyses of gaze, gesture, and voice.

[3] For a detailed description of the synthesis rule see Cappella (1980).

[4] One possible objection to the findings is the limited sample size and narrow time window (sampling at 0.1 seconds.). A structurally similar transition matrix based on 40 dyads of various types and a sampling interval of 0.3 seconds shows the counter diagonal probabilities with the same pattern as in table 2. They are all near zero, confirming the claim that there is mutuality and negotiation in changing smiling states for people in cooperative interaction (data for this matrix can be seen in Cappella, 1993 or are available from the author by request).