Relational Agents: 
Effecting Change through Human-Computer Relationships

by
Timothy Wallace Bickmore

Master of Science, Computer Science, 
Arizona State University, 1986
Bachelor of Science and Engineering, Computer Systems Engineering 
Arizona State University, 1985

Submitted to the Program in Media Arts and Sciences, 
School of Architecture and Planning, 
In partial fulfillment of the requirements for the degree of

Doctor of Philosophy 
at the 
Massachusetts Institute of Technology 
February 2003

© 2003 Massachusetts Institute of Technology, All rights reserved.

Author  January 31st, 2003

Certified by  Rosalind W. Picard
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Certified by  Justine Cassell
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by  Andrew B. Lippman
Chairperson, Departmental Committee on Graduate Students
Chapter 3

THEORY

Although relational agents could be constructed without reliance on a deep theory of what a social-emotional relationship is, such a theory can provide a basis for the necessary flexibility and generality required for an agent that can function adaptively over a wide range of contexts, including multiple interactions over long time spans. In addition, a theory of human-computer relationships can provide a framework within which the design and evaluation of relational agents can be better understood. This chapter integrates work from social psychology, linguistics, communication and artificial intelligence to provide such a theoretical framework.

I will start by reviewing fundamental definitions of personal relationship from the field of social psychology, then discuss artificial intelligence theories of multi-agent collaboration and accommodation that will be used as the basis for a theoretical framework before presenting the synthesized theory of what a relationship is. I then proceed to discuss how this theoretical framework can be applied to understanding aspects of relational behavior in two realms: the micro-structure of face-to-face conversation and the macro-structure of relational maintenance in long-term relationships.

3.1 The Social Psychology of Personal Relationships

Most recent work in the social psychology of personal relationships takes a fundamentally dyadic approach to the concept of “relationship” (Berscheid & Reis, 1998). Kelley et al define this concept as referring to two people whose behavior is interdependent, in that a change in the state of one will produce a change in the state of the other (Kelley, 1983). Thus, a relationship does not reside in either partner alone, but in their interaction with each other. Further, a relationship is not defined by generic patterns of interaction (e.g., associated with stereotypical roles), but by the unique patterns of interaction for a particular dyad (Berscheid & Reis, 1998).

In addition to defining relationships in terms of what people actually do together and their degree of interdependence, relationships are also often defined in terms of what the people in them provide for one another. Duck, for example, defines the following list of provisions that “friends” in our culture are expected to provide for each other (Duck, 1991):

- Belonging and a sense of “reliable alliance”. The existence of a bond that can be trusted to be there for a partner when they need it.
- Emotional integration and stability. Friendships provide necessary anchor points for opinions, beliefs and emotional responses.
- Opportunities for each partner to talk about themselves. Friendships help fulfill the need for self-expression and self-disclosure.
- Provision of physical, psychological and emotional support. Physical support involves doing favors, such as giving someone a ride or washing the dishes. Psychological support involves showing appreciation for the other and letting them
know their opinions are valued. Emotional support includes affection, attachment and intimacy.

- Reassurance of worth and value, and an opportunity to help others. We value friends because of their contribution to our self-evaluation and self-esteem, directly via compliments and indirectly by telling us of the good opinions of others. Also, friends increase our self-esteem by simply attending to us, by listening, asking our advice and valuing our opinions.

Similarly, Brehm defines “intimate relationships” in terms of provisions such as intimacy, nurturance, assistance, social integration, and reassurance (Brehm, 1992).

This notion of provisions is also reflected in economic models of relationship, such as exchange theory (Brehm, 1992), to the extent that they model relationships in terms of costs vs. benefits. Social exchange models are economic models of the costs, benefits, investments, and alternatives individuals have in relationships and how these relate to their levels of commitment. Social exchange models have received more empirical validation than any other theoretical framework in the social psychology of personal relationships. In these models the benefits of a relationship can be seen as the perceived value of the provisions that one partner receives while their costs can be seen as the perceived cost of the provisions that they provide to the other. In these theories a relationship only exists when there is such an economic exchange, and the longevity of the relationship (commitment to continue) can be reliably predicted from equations involving these costs and benefits and other terms such as perceived alternatives to and amount of investment in the relationship.

Many researchers have also attempted to define dimensional models that identify the underlying features that characterize different stereotypical relationships. The most commonly mentioned dimensions are power and social distance, with social distance further refined by many researchers into as many as 14 sub-dimensions (Brown & Gilman, 1972; Burgoon & Hale, 1984; Spencer-Oatey, 1996; Svennevig, 1999). Other dimensions used to characterize relationships include equal vs. unequal, hostile vs. friendly, superficial vs. intense, informal vs. formal (Wish, Deutsch, & Kaplan, 1976), and various typologies of love (Brehm, 1992).

Some work has also been done on integrating these various models. For example McGuire derived a set of 72 types of helping behavior and grouped them into four factors—casual helping (low cost to the helper), substantial personal helping (substantial cost to the helper), emotional helping, and emergency helping—and then determined how the frequency of these varied with the interpersonal closeness of the relationship between the helper and helpee, effectively bridging provisional and dimensional models of relationship (McGuire, 1994).

### 3.2 Multi-Agent Collaboration

Relationships involve collaborative behavior; coordinated activity in which the participants work jointly with each other to satisfy a shared goal. Examples of such collaborations involve coordination on specific activities within a relationship (e.g., washing and drying the dishes, reminiscing) as well as collaboration on the relationship itself (e.g., negotiating roles). To begin to formalize a notion of relationship for relational agents, I now turn to the field of artificial intelligence and formalisms for planning and collaboration among autonomous agents.
Perhaps the most complete theory of multi-agent collaboration in the field of AI is the SharedPlans theory developed by Grosz & Sidner, et al. (Grosz & Kraus, 1993; B. Grosz & S. Kraus, 1996; Grosz & Sidner, 1990). This work describes a normative, performance model detailing how agents move from individual goals and intentions into collaborative, coordinated activity, based on representations in the minds of the individual agents.

The SharedPlans theory is based on the theories of Bratman and Pollack, who outline a mental-state view of plans in which having a plan is not just knowing how to do an action, but also having the intention to do the actions entailed (Bratman, 1990; Pollack, 1990). This approach thus differentiates between knowing how to accomplish a goal (a "recipe") and having a plan, which includes intentions. The SharedPlans theory explains how a group of agents can incrementally form and execute a SharedPlan that then guides and coordinates their activity towards the accomplishment of a shared goal.

Informally, two agents are said to have a SharedPlan when they mutually believe that: a) they have a common goal; b) they have agreed on a recipe to accomplish the goal; c) they are each capable of performing their assigned actions; d) each intends to do their assigned actions; and e) they are committed to the overall success of the collaboration. SharedPlans are usually incrementally refined and executed by the collaborating agents; in a typical scenario a SharedPlan is initially partial (incompletely specified) and only becomes completely specified once the agents have finished refining and executing it. Refinement of a partial plan is carried out through means-ends reasoning and negotiation among the agents.

The formalization presented in (B. J. Grosz & S. Kraus, 1996) uses a first-order logic augmented with several modal operators, meta-predicates, and action expressions. Table 3-1 lists the notations from that work that will be used here (I am abstracting from their representation; reference to the time, context and plan identifier parameters are not needed for the current discussion and have been omitted). In this representation, actions can be readily executable (“basic-level”) or complex, with complex actions having recipes consisting of additional sub-actions that are basic-level or complex. Thus actions and their decomposition form recipe trees, which reflect hierarchical plan decomposition.

The theories rely on two different notions of intention. Both notions follow Bratman in that they prevent the agent(s) from adopting conflicting intentions. “Intend to” (Int.To) is an attitude a single agent holds with respect to an action, requires that the agent know a way of doing the action (e.g., a recipe) and commits it to means-end reasoning, if required, to carry the action out. An agent cannot intend for another agent to perform an action using this attitude. “Intend that” (Int.Th) is an attitude one or more agents hold with respect to a proposition, and reflects the desire for the agent(s) to achieve the state of the world specified, however, it is not as strong a commitment as intend-to, and does not require that the agent(s) know how to achieve the action and does not commit them to means-end reasoning. Further, an agent can intend-that another agent achieve the specified proposition.
<table>
<thead>
<tr>
<th>Type</th>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literals</td>
<td>G1, G2</td>
<td>Agents</td>
</tr>
<tr>
<td></td>
<td>GR</td>
<td>a group (set) of agents</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>a state of the world (a proposition)</td>
</tr>
<tr>
<td></td>
<td>(\alpha)</td>
<td>an action</td>
</tr>
<tr>
<td></td>
<td>(R_{\alpha})</td>
<td>a recipe to achieve (\alpha)</td>
</tr>
<tr>
<td>Modal Operators</td>
<td>Bel(G,P)</td>
<td>Belief</td>
</tr>
<tr>
<td></td>
<td>MB(GR,P)</td>
<td>mutual belief</td>
</tr>
<tr>
<td></td>
<td>Int.To(G, (\alpha))</td>
<td>Agent intends to perform (\alpha).</td>
</tr>
<tr>
<td></td>
<td>Int.Th(GR,P)</td>
<td>Agent(s) intend that state P be achieved.</td>
</tr>
<tr>
<td></td>
<td>Do(GR,(\alpha))</td>
<td>Agent(s) G do (\alpha).</td>
</tr>
<tr>
<td>Meta-predicates</td>
<td>FSP(GR, (\alpha), (R_{\alpha}))</td>
<td>Agents have a full shared plan to do (\alpha) using (R_{\alpha}).</td>
</tr>
<tr>
<td></td>
<td>PSP(GR, (\alpha))</td>
<td>Agents have a partial shared plan to do (\alpha).</td>
</tr>
<tr>
<td></td>
<td>SP(GR,(\alpha))</td>
<td>The group has a FSP or a PSP and a FSP to elaborate it.</td>
</tr>
<tr>
<td></td>
<td>CBAG(GR,(\alpha), (R_{\alpha}))</td>
<td>A group of agents “can bring about” (\alpha) using (R_{\alpha}).</td>
</tr>
</tbody>
</table>

**Table 3-1. SharedPlans Notations Used (B. J. Grosz & S. Kraus, 1996)**

Grosz and Kraus define a Full Shared Plan as a meta-predicate that represents the situation in which two (or more) agents have determined the recipe by which they are going to achieve some action, and the members of the group have adopted intention-to toward all of the basic-level actions in the recipe as well as intentions-that toward the actions of the group and its other members. A Full Shared Plan is represented as \(\text{FSP}(\text{GR},\alpha,R_{\alpha})\), where: GR is the group of agents involved in the collaboration; \(\alpha\) is the action to be achieved, and \(R_{\alpha}\) is the recipe to be used. A Partial Shared Plan—\(\text{PSP}(\text{GR},\alpha)\)—differs from a Full Shared Plan in that the agents may only have a partial recipe for doing the action, they may have only partial plans for doing some of the steps in the recipe, and there may be some sub-actions which the have not yet been assigned to a particular agent.

While relationships certainly involve collaborative behavior, it is worth pointing out at this point that a collaboration, as modeled by a SharedPlan, is not a good representation for the concept of “having a personal relationship”. Although not fully specified here, SharedPlans involve the accomplishment of a specific goal, using specific actions over a specific time interval. Relationships, on the other hand, are typically unbounded in duration, and while the range of activities conducted within the relationship can be specified (e.g., as provisions) the particular activities that a dyad engages in at any one time cannot be defined. Further, while specific actions may be required to build, change, maintain or terminate a relationship, no actions are required to simply “have” a relationship (e.g., partners can say they’re friends even if they haven’t talked to each other in ten years).

Grosz and Sidner discuss general strategies for moving from one agent’s having a goal to a group of agents having a SharedPlan to achieve that goal (Grosz & Sidner, 1990). They bridge part of this gap via a “conversational default rule” (a rule which operates in the
absence of evidence to the contrary) which states that if one of the agents has a goal for P and they are cooperative and communicating about this desire to achieve P, then they will come to mutually believe that this agent has a desire the group to construct a SharedPlan to achieve P. However, to move from this mutual belief about G1’s desire for a collaboration to actually performing the collaboration requires (probable) negotiation and assent by G2. How can agents assess whether their potential collaborators (such as G2) will be likely to provide assistance? These expectations of future collaborations can be defined by appeal to the notion of accommodation, which I turn to next.

3.3 Accommodation

Accommodation is the situation in which one agent infers the goals of a second and takes action to help without the first agent making an explicit request. It is hypothesized to underlie a wide range of collaborative and linguistic behavior. Thomason defines accommodation as follows:

Most generally, accommodation consists in acting to remove obstacles to the achievement of desires or goals that we attribute to others... I am accommodating you, for instance, if I open the door when I see you approach it with your hands full of packages. (Thomason, 1990)

This notion can be directly encoded in the SharedPlans formalism above as the following meta-predicate, representing the situation when G2 accommodates goal P for G1:

\[
\text{Acc1}(G1,G2,P) \Leftrightarrow \\
(1) \quad \text{Bel}(G2,\text{Int.Th}(G1,P)) \land \neg \text{Bel}(G2,P) \land \\
(2) \quad \text{Bel}(G2,(\exists R\alpha)[\text{Do}(\{G1,G2\},\alpha) \Rightarrow P] \land \text{CBAG}(\{G1,G2\},\alpha,R\alpha)]) \\
(3) \quad \Rightarrow \text{Int.Th}(G2,\text{SP}(\{G1,G2\},\alpha))
\]

That is, whenever (1) G2 believes that G1 desires to achieve P (and believes that P is not already satisfied) and (2) believes it is able to help (believes there is a recipe whose execution will result in P and which the agents are able to collaborate on in order to achieve P) then (3) G2 will adopt the intention to collaborate with G1 on P’s achievement.

Using this notion of accommodation, while a good starting place, has several shortcomings if it is to be used as the basis for a provision-based definition of relationship. First, in order for agents to be able to ‘keep score’ of when other agents have actually helped them in the past, accommodation needs to be based on mutual knowledge, rather than just on beliefs in G2’s head (i.e., G2 doesn’t get ‘credit’ for helping unless everyone is aware that helping is going on). Second, relationships are not only defined by the goals each partner will help the other with, but the manner in which these goals are satisfied; namely by the set of recipes that the partners will use to help the other. That is, G2 may not be willing to use any recipe at its disposal in helping G1 with P, but only a specific subset of those it knows about. Further, it may be characteristic of this dyad for them to use a particular recipe together when performing a particular kind of task.

A second definition of accommodation that captures these additional subtleties is:
In this definition, (1) both agents must believe there is a specific recipe that will bring about P, and (2) when there is mutual belief that G1 desires P (and P is not already satisfied), and (3) they can achieve P by collaborating on the execution of the recipe, then (4) G2 will adopt the intention to collaborate with G1 on P’s achievement. Note that the condition in (3) lets G2 “off the hook” for accommodating if it is presently unable to help out.

Essentially, this is an expectation, a future potential for action. The inner MB indicates that it’s not enough for G2 to simply observe that G1 desires P (e.g., through observation or third parties) and act in secrecy to aid G2. In relational accommodation both agents need to be aware that accommodation is happening. As mentioned above, both agents are very interested in 'keeping score' of when others actually accommodate them (an important part of relationships, for example, as represented in exchange models), and this can't happen unless everyone is aware that accommodation is intentionally going on. The outer MB allows the agents to plan their lives in such a way that they can rely on the accommodation of others. It turns out that line (4) is actually too strong, but it's the ideal case. In reality, there is a great deal of uncertainty about what one agent will help another one with, and agents constantly assess the strength of these accommodation relations.

3.4 A Formal Definition of Relationship

Given the definition of relational accommodation above, we can proceed to define a relationship between two agents as the set of all such accommodation relations that hold between them:

\[
\text{Relationship}(G1,G2) \iff \{ <x,y,P,R_\alpha> \mid x,y \in \{G1,G2\} \& R_\alpha \in \text{Recipes} \& \text{Acc2}(x,y,P,R_\alpha) \}
\]

Two agents are said to 'have a relationship' if this set is non-empty.

Under this definition, "having a relationship" or "being in a relationship" is not a collaboration (SharedPlan), but is a set of expectations for future collaborations should the need arise. On the other hand, "establishing a relationship", "changing a relationship", "maintaining a relationship" are specific goals which require collaboration (and SharedPlans) in order be accomplished, at the time they are performed.

As mentioned above, the normal state of most human relationships (excluding formalized ones such as contractual relationships) is that there is a great deal of uncertainty regarding which accommodation relations actually hold. That is, an agent is constantly trying to update its degree of belief in each of the things that its partner will accommodate it on, as well as its own likelihood of accommodating its partner on a variety of activities. It is these assessments, and the desire for information about them, that drives most relational behavior in human social interaction. And, it is the agreement on these assessments between partners and their
degree of overlap with stereotypical relationships that ultimately defines the relationship between them.

Following the observations from the social psychology of personal relationships in section 3.1, the particular activities that two individuals expect to engage in with each other are (potentially) unique to that dyad, however to the extent that this set of activities overlaps with a common relational stereotype (e.g., "friend", "colleague", "acquaintance", etc.), then the dyad can be said to be in a relationship of that type, governed by the extent of the overlap.

While a dyad’s interaction history plays a strong role in establishing their relationship (the set of accommodation behaviors they will readily engage in), and provides content for continuity behaviors (referring back to shared experiences) and other functions, their history is not an intrinsic part of the representation of their relationship according to the above definition. To see why, note that dyads with relational expectations but no history (e.g. couples in an arranged marriage, business people in contractual business arrangements, etc.) are said to be in a relationship, at least from the first moment of acting in accordance to their relational expectations, whereas dyads with a history but who have no further expectations of relational interactions (e.g., deceased partner, real estate agent following a closing, divorcees, etc.) are said to no longer be in a relationship.

Finally, while the definition above states explicitly what two agents will do for each other, it also implicitly defines what they will not do with each other. If there is no expectation that G2 will accommodate G1 on $\alpha$, then G1’s proposing a collaboration on $\alpha$ is “marked” and must be carefully negotiated (discussed further in 3.7). This view allows common relational maxims such as “avoid intimate topics”, “avoid impositions” or “avoid face threats” to be expressed by the exclusion of activities that would violate these rules from the set of activities sanctioned by the relationship.

3.5 Keeping Score

It is in an agent’s best interest to know what it can rely on its relational partners for, so that it can plan its life accordingly. Unfortunately, except in certain formal situations (e.g., contractual relationships) this information has to be inferred because relationships are typically in a state of flux, and because rejections (finding out that a partner won’t accommodate you on a specific activity) are significant blows to one’s self esteem and represent “face threats” (Goffman, 1967).

An exact calculus of relational expectations has yet to be developed, but there are some general psycho-social principles that can be used. The most fundamental principle is that when a dyad engages in a particular collaborative activity, it increases each of their expectations that they can engage in the same activity again in the future, absent explicit evidence to the contrary. This implies that one way to change a relationship is to simply do something new together. This also implies that an existing relationship can be maintained (expectations kept high) simply through periodic performance of the collaborative activities that constitute the relationship.

Partners also perform explicit tests of the status of their relationship by proposing activities just to see if they get uptake or not (e.g., “why don’t you come meet my parents?”) or to simply confirm what they think the current status of the relationship is (Duck, 1991).

Relational stereotypes also play a significant role in relational expectations. To the extent that the set of accommodation behaviors defining a dyad’s relationship overlaps with a socio-
cultural relational stereotype, their expectation of being able to perform any of the recipes indexed by the stereotype is increased. Thus, as a dyad begins to perform activities together, their expectations about continuing those activities and similar activities normally increase, where similarity here is partly governed by the stereotypes. For example, once a dyad begins to do many of the things that "friends" do (in a particular culture) they can expect to do all of the other things friends normally do together, absent reasons for thinking otherwise.

In addition to actual past history and relational stereotypes, expectations can be influenced by generalized notions of trust in one’s relational partner.

3.5.1 Trust

Although agents may keep track of separate relational expectations for the entire range of accommodation behaviors they expect to collaborate on with their partner, it is useful (e.g., for planning purposes) to have generalized expectations about the likelihood of a partner meeting one’s relational expectations. Such generalized expectations are called trust and, in social psychology, are generalized over all interactions with an individual ("specific interpersonal trust" (Johnson-George & Swap, 1982)), over all interpersonal relationships ("generalized trust", strongly related to attachment style (Berscheid & Reis, 1998)), or over all interactions in a particular context (e.g., professional vs. social relationships (Barber, 1983)).

Trust is a central concept in relationships, both because it is a prerequisite for any kind of collaborative activity, and because it is one of the primary outcomes of and reasons for engaging in relationship development. The literature on trust spans the disciplines of sociology, social psychology, and philosophy, and several computational models of trust have been developed in the software agents and distributed AI fields as well.

Relationally, trust is an antecedent to self-disclosure (Wheless & Grotz, 1977) and is an important component of intimacy (Berscheid & Reis, 1998). Trustworthy sources are also generally more persuasive (at least for certain types of decisions) (Petty & Wegener, 1998).

Deutsch provides a detailed theory of trust, validated through a series of empirical studies of cooperation among human subjects (Deutsch, 1973). In his theory he suggests that trust involves an analysis of the benefits vs. the potential costs to the trustor for taking a particular action. He suggests that trust comes into play in those situations in which one makes a decision whose negative consequences potentially outweigh the positive consequences. He proposes a utility theoretic framework in which a trusting decision is made when the outcome of that decision is an ambiguous situation in which the following holds:

\[ V_a^+ < V_a^- \]

and

\[ V_a^+ \times S.P. > V_a^- \times S.P. + K \]

Where,

\[ V_a^+ \] is an event having positive motivational significance (value, or benefit).
\[ V_a^- \] is an event having a negative motivational significance (value, or benefit).
\[ S.P. \] is the subjective probability of attaining \[ V_a^+ \].
\[ S.P. \] is the subjective probability of \[ V_a^- \].
\[ K \] is the "security level" that the individual needs for action (trait).

The subjective probabilities involved in this assessment are based on many factors, but the most important ones are one’s own past experiences in similar situations, the past experiences of others, the opinions held by others whom one respects, one’s personal
assumptions about the benevolence-malevolence of the reality one is in, and one’s confidence about being able to influence the occurrence of Va+ or the nonoccurrence of Va through one’s own actions or through available help. As an example, suppose I asked a close friend to mail an important letter for me. This is a trusting action on my part, because Va+ < Va−, where Va− is the cost of the letter not getting mailed, but my assessment of S.P.− is sufficiently low with respect to my assessment of S.P.+ that I decide to take this action anyway.

One of the most comprehensive computational models of trust was developed by Marsh to research models of cooperation among agents in Distributed Artificial Intelligence simulations (Marsh, 1994). Marsh investigated the role of trust in the initiation and maintenance of cooperation between agents. In his framework an agent x decides to cooperate with another agent y in situation α when:

$$T_x(y,α) > \text{Cooperation\_Threshold}_x(α)$$

Where, $T_x(y,α)$ is x’s trust in agent y for situation α, and is given by:

$$T_x(y,α) = \left[U_x(α) + \hat{T}_x(y)\right] \times I_x(α) \times \left(\hat{r}_y(x)\right)$$

And,

$U_x(α)$ is the utility of α to x.

$\hat{T}_x(y)$ is x’s general trust in y.

$I_x(α)$ is the importance of α to x.

$\left(\hat{r}_y(x)\right)$ is x’s estimate of y’s trust in x.

The two estimates of trust are computed by integrating over past experience with the other agent (initialized with a "generalized" trust value). Marsh presents three simple ways of doing this--optimistic estimation (maximum), pessimistic estimation (minimum), and "pragmatic" estimation (averaged)--and also a method which takes into account a fixed memory of agents by averaging over a window of the last N interactions.

The Cooperation Threshold is defined as:

$$\text{Cooperation\_Threshold}_x(α) = \frac{\text{Perceived\_Risk}_x(α)}{\text{Perceived\_Competence}_x(y,α)} \times I_x(α)$$

Where, Perceived_Risk is x’s perception of the risk associated with situation α, and Perceived_Competence is x’s perception of y’s competence in performing in situation α.

While the exact form of Marsh’s model is debatable, he has motivated his choice of variables and their direction of contribution in the estimation of trust and cooperation from extensive research into the literature on trust. It seems clear that trust plays a role in an agent’s assessments of whether another agent will accommodate it or not, or whether it should accommodate another agent, and that the estimation of trust involves a cost/benefit analysis involving reasoning about past experiences with the other agent, the importance or utility of the outcome, and the risks involved. One drawback of using Marsh’s work for a relational model is that it provides a generalized expectation; a calculus for determining the influence of
this on the specific expectations that an agent has for accommodation has yet to be determined.

3.6 Motivation to Meet Relational Expectations

Why should agents bother to act in accordance with the relational expectations they have established with their partner? One reason is to simply maintain the relationship so as to not lose any benefits that might later accrue from it. Many relationships, such as friendships, are voluntary, and violations of the “unwritten contract” between the partners can result in relationship dissolution (Wiseman, 1986), resulting in loss of provisions to both partners.

3.6.1 Rights & Obligations

Another motivation for not violating relational expectations is the threat of retaliatory action by one’s partner and the society in general. These actions are sanctioned under the concepts of rights and obligations.

As Jackendoff observed, rights and obligations are fundamental to the fabric of human social organization and appear to be universal concepts in human societies (Jackendoff, 1999). Jackendoff’s analysis is directed at the question of how people conceptualize situations in which someone can be said to have a right or an obligation, and relies on the formal framework of "conceptual semantics", which encodes regularities in people's minds (rather than, e.g., formal statements about the real world). In this framework, rights and obligations are like facts-- "objectively determinable persistent entities"--that have their own logic which Jackendoff guesses might possibly be innate (aside from their universality, even primates seem to have some notion of a moral dimension of obligations). Rights and obligations can be either persistent and universal with respect to some class of actions, or existential in which, like intentions, once satisfied (an obligation fulfilled or a right exercised) they cease to exist. Rights have an inherently positive value, and are thus to be desired by an agent, whereas obligations have an inherently negative value.

A fundamental aspect of rights and obligations that distinguishes them from other kinds of modals (such as may and should) is the consequence of noncompliance. The inference rule for obligation noncompliance can be paraphrased as:

IF G2 has an obligation to do an action $\alpha$ for the benefit of G1 AND
G2 does not perform this action (in a suitable time-frame)
THEN
G1 has the (existential) right to do an action $\beta$ that will have negative value for G2 in exchange for G2's noncompliance

The parallel inference for rights is:

IF G1 has a right to perform an action $\alpha$ AND
G2 prevents G1 from performing $\alpha$
THEN
G1 has the (existential) right to do an action $\beta$ that will have negative value for G2 in exchange for G2's violation
There are also moral dimensions to these phenomenon. In the case of unfulfilled obligations, all members of the society are morally justified in sanctioning someone for their noncompliance, whereas in the case of violated rights, everyone is morally justified in sanctioning the offender.

Traum also introduced an attitude of obligation in his work on modeling intentions in dialogue (Traum & Allen, 1994). In his model, an agent’s behavior is determined by a number of factors, including the agent’s current goals in the domain, and a set of obligations that are induced by a set of social conventions. The agent considers both goals and obligations, to the extent that it can, when planning actions. He used this framework to explain how and why an agent responds to dialogue moves by stating that certain moves obligate the listener (relative to some set of social norms) to respond. For example, when asked a question this creates an obligation in the listener to respond with an answer.

3.6.2 Implications

In light of Jackendoff’s analysis, relational expectations of accommodation can be seen as persistent obligations to the agent who is to be accommodating, and persistent rights to the agent who expects to be accommodated. In this view, violating relational expectations opens up the possibility of retaliation by the thwarted partner and, especially to the extent that the relationship fits a socio-cultural stereotype, by other members of society. Relative to Traum’s model, relational expectations actually fit more in line with his view of “social norms” in that they are persistent and give rise to specific collaborative behavior when the need arises.

3.7 Relational Dynamics

As defined above, a relationship not only defines what a dyad can do together, but implies that any activities not sanctioned by the relationship are forbidden. Given that most relationships are continuously evolving, how do changes in relational expectations occur? They must be carefully negotiated. Explicit forms of negotiation can be, in themselves, very subtle and complex. However, relational negotiation is even more complex, since it is usually conducted in a tacit, off-record manner, with the bids, uptakes, and rejections handled in an indirect manner to prevent explicit rejection and loss of self-esteem.

3.7.1 Negotiation

The basic form of negotiation occurs when one agent makes a proposal for a new activity and the other agent accepts or rejects it. Human relational negotiations are typically very subtle and much more complex. However, even when negotiation moves are entirely explicit, negotiation is still a very complex phenomenon. Sidner defines an artificial language within which negotiations can take place, and includes actions such as AcknowledgeReceipt (of a proposal), Counter (a counter-proposal), and RetractProposal in addition to the basic propose and accept actions (Sidner, 1994).

3.7.2 Face

One theory that sheds a lot of light on the forms in which relational negotiation must occur is Goffman’s notion of “face”. This construct came out of work in sociology (Goffman, 1967) and has been most productively employed in the field of sociolinguistics (Brown & Levinson, 1987). It has been shown to govern a significant amount of behavior in social
interactions, and is itself governed by the nature of the relationship between the interactants. "Facework" (strategies to manipulate this construct) can be used to change the nature of a relationship, and is thus relevant to the design of relational agents.

In Goffman’s dramaturgical approach to social interaction, he defined an interactant’s "line" as the patterns of action by which an individual in an interaction presents an image of himself and the situation, that is his social role in the current joint activity (Goffman, 1967). The notion of "face", Goffman went on to say, is "the positive social value a person effectively claims for himself by the line others assume he has taken during a particular contact". An interactant maintains face by having their line accepted and acknowledged by their interactants. Events which are incompatible with their line are "face threats" and are mitigated by various corrective measures if they are not to lose face.

Brown and Levinson used Goffman’s notion of face in the formulation of their theory of the use of politeness forms in language (Brown & Levinson, 1987). They defined positive face as an individual’s desire to be held in esteem by their interactants, and negative face as an individual’s desire for autonomy, and went on to characterize the degree of face threat of a given speech act as a function of power, social distance, and the intrinsic threat (imposition) imposed by the speech act. Brown and Levinson’s theory accounts for some cross-cultural universals in sociolinguistic behavior by explaining when individuals are likely to use different forms of politeness in making a request, given the nature of the request and the relationship between the interactants.

Relative to my theory of relationship, facework can be seen as predicting the kind of mitigating action required when a change in relationship is attempted. When actions sanctioned by the relationship are performed, no mitigating action is required (and, via accommodation, no negotiation of any kind need occur). However, when actions outside the relationship are proposed, different kinds of mitigation are called for, from conventional politeness to going “off record” with a request and thereby making it very easy for the helper to reject the proposal without threatening the requester’s self-esteem. The degree to which the action is unexpected, relative to the existing relationship, determines the form of mitigation required (per Brown & Levinson) and the dimensions along which this degree of unexpectedness are measured are, according to Brown & Levinson, power and solidarity, the two primary dimensions used in dimensional models of relationship.

Facework normally plays out in the micro-structure of face-to-face conversation. Examples of strategies that can be employed include: hedged or indirect requests ("You wouldn't possibly want to go to the movies, would you?"); pre-requests ("Do you like movies?"); pre-invitations ("What are you doing this evening?"); and pre-announcements ("You know what I'd like to do?") (Levinson, 1983). Rejections are almost always indirect and often nonverbal, including such behaviors as pausing (allowing the proposer to retract their suggestion), gazing away, preface markers ("Uh", "Well"), and affective facial displays (rejections are a type of "dispreferred" conversational action (Levinson, 1983)).

### 3.7.3 Other Forms of Change

Another method of changing relational expectations is for one partner to simply start acting as if the change had already occurred. For example, Lim posits that when an individual is attempting to increase closeness in a relationship they will perform less than expected facework for "normal and familiar" joint activities, while those who are trying to increase
distance in a relationship will use facework strategies normally used in more distant relationships (Lim).

When one partner starts acting as if a change had occurred and the other goes along with it, the change behavior itself is a form of accommodation, and may lead both partners to believe that this method of change can be used again in the future.

3.8 The Affective Dimension

Much has been said about social dimensions of relationship, but not much about the emotional aspects thus far. However, emotions play a crucial role in all human relationships. Relative to the relational theory presented above, emotions (e.g., love, happiness), their grounding (through empathy), and management (through emotional support) are fundamental provisions in most close personal relationships. Further, deviations from relational expectations give rise to some of the strongest emotions humans feel (e.g., shame, guilt, embarrassment, jealousy, and social anxiety) and thus serve a regulatory role in helping to ensure that partners meet each other’s expectations.

3.8.1 Affective Relational Provisions

Emotions (e.g., love, joy) and behaviors that help manage emotions (e.g., empathy) are among the most important provisions of close personal relationships. Many kinds of relationship stereotypes, such as passionate love, are defined in terms of the emotions felt by one partner towards the other (Brehm, 1992).

According to Goleman, being able to manage emotions in someone else is the core of the art of handling relationships (Goleman, 1995), and the ability to provide emotional support is also frequently mentioned as one of the requisite characteristics of a good friend or intimate partner (Cole & Bradac, 1996; Gill, Christensen, & Fincham, 1999; Goldsmith, McDermoot, & Alexander, 2000; McGuire, 1994). Emotional communication between partners is a crucial ingredient in most relationships. Relationship quality is often defined in terms of the quality of emotional communication between partners; "it is taken for granted that marital harmony is strongly related to effective communication between spouses, and that effective communication, to some optimal degree, involves spouses' understanding of each others' thoughts and feelings" (Berscheid & Reis, 1998). Empathy—the process of attending to, understanding, and responding to another person's expressions of emotion—is a pre-requisite for providing emotional support which, in turn, provides "the foundation for relationship-enhancing behaviors, including accommodation, social support, intimacy, and effective communication and problem solving" (Berscheid & Reis, 1998) (Okun, 1997).

In addition to being important in intimate relationships, emotional support provisions are also crucially important in most helping professions. The most significant empirical support of this phenomenon is in the field of psychotherapy, in which measures of “working alliance”—the trust and belief that the therapist and patient have in each other as team-members in achieving a desired outcome—show consistently high correlations with successful outcomes (Horvath & Symonds, 1991). Even in physician-patient interactions, physician empathy for a patient plays a significant role in prescription compliance, and a physician’s lack of empathy for a patient is the single most frequent source of complaints (Frankel, 1995).
3.8.2 The Regulatory Role of Affect

Relationships play a central role in the emotional life of all humans. According to Lazarus "most emotions involve two people who are experiencing either a transient or stable interpersonal relationship of significance" (Lazarus, 1994), and Bowlby contends that most intense emotions arise when people are forming, maintaining, disrupting, terminating, or renewing close relational ties with others (Bowlby, 1979). Many emotions, such as jealousy or passionate love, can only occur within a relationship, while other emotions, such as loneliness, occur because of problems with or lack of a desired social relationship.

One function of these strong emotions is to serve as basic regulatory mechanisms to ensure that relational expectations are met. Violations of relational expectations give rise to emotions such as shame, guilt, embarrassment, and social anxiety and act as a motivator for the partner feeling them to take mitigating action.

3.8.3 Affect in Relational Negotiation

Emotions also play an important role in the negotiation process of relationship building and maintenance. As stated above, relational negotiation is usually conducted in a tacit, off-record manner, with the proposals and rejections handled in an indirect manner to save face. Uptakes and rejections, in particular, are often achieved through positively and negatively valenced emotional displays (e.g., happiness display for uptake, disgust, contempt or anger displays for rejection).

3.9 Application to the Micro-Structure of Face-to-Face Conversation

Face-to-face conversation is the primary modality used to build and maintain human relationships. It should be apparent from the definitions above that negotiation of relational expectations, for most types of relational provisions, would be very difficult to accomplish without language. According to Duck, “Language—or more broadly, communication and all that is culturally encoded within it—is … a crucial basis for establishing ways in which we conduct human relationships and judge their quality” (Duck, 1998). In other work, Duck even defines relationships in terms of "shared meaning systems" whereby the individuals in the relationship come to see and evaluate events in the same way, principally through the use of language (Duck, 1995).

Language is also required in order to enact many kinds of relational provisions, such as informational support, and it seems to be especially important in emotional support processes. For example, there is much evidence that talking about traumatic events helps people to deal with them. According to Burleson and Goldsmith, all of the emotions, images and thoughts associated with a trauma are held together by the traumatic incident itself, and talking about the event may help dismantle the phenomenal whole that traumas constitute (Burleson & Goldsmith, 1998). The very process of putting the event into words, subject to the syntactic and pragmatic constraints of language and the need to decontextualize it to describe it to another, helps the distressed person get a new perspective on the event and reappraise it. Another series of studies showed that when people fail to articulate a traumatic event verbally, they fail to deal with it effectively. Thus, conversation seems to play a central role in coping with emotion. (Andersen & Guerrero, 1998b).

Not all forms of language use, however, are equally effective for relational communication. Face to face conversation is the primary, universal mode of human-human
communication. According to Fillmore, "the language of face-to-face conversation is the basic and primary use of language, all others being best described in terms of their manner of deviation from that base" (Fillmore, 1975), and Levinson,"face-to-face interaction is not only the context for language acquisition, but the only significant kind of language use in many of the world’s communities, and indeed until relatively recently in all of them" (Levinson, 1983). Face-to-face conversation is particularly important for relational communication because of the availability of nonverbal cues, such as facial expressions, body posture and hand gesture.

3.9.1 The Role of Nonverbal Behavior

According to Argyle, nonverbal behavior is used to express emotions, to communicate interpersonal attitudes, to accompany and support speech, for self presentation, and to engage in rituals such as greetings (Argyle, 1988). Next to coverbal behavior and emotional displays, the most important use of nonverbal behavior in relational dialogue is the display of interpersonal attitude. The display of positive or negative attitude can greatly influence whether we approach someone or not and our initial perceptions of them if we do.

The most consistent finding in this area is that the use of nonverbal "immediacy behaviors"--close conversational distance, direct body and facial orientation, forward lean, increased and direct gaze, smiling, pleasant facial expressions and facial animation in general, nodding, frequent gesturing and postural openness--projects liking for the other and engagement in the interaction, and is correlated with increased solidarity (perception of “likemindedness”) (Argyle, 1988; Richmond & McCroskey, 1995). Other nonverbal aspects of "warmth" include kinesic behaviors such as head tilts, bodily relaxation, lack of random movement, open body positions, and postural mirroring and vocalic behaviors such as more variation in pitch, amplitude, duration and tempo, reinforcing interjections such as "uh-huh" and "mm-hmmm", greater fluency, warmth, pleasantness, expressiveness, and clarity and smoother turn-taking (Andersen & Guerrero, 1998a).

The importance of nonverbal behavior for relational communication is also supported by studies comparing the affordances of different communication media. Several studies have found that the additional nonverbal cues provided by video-mediated communication do not affect performance in task-oriented interactions, but in interactions of a more relational nature, such as getting acquainted, video is superior (Whittaker & O'Conaill, 1997). These studies have found that for social tasks, interactions were more personalized, less argumentative and more polite when conducted via video-mediated communication, that participants believed video-mediated (and face-to-face) communication was superior, and that groups conversing using video-mediated communication tended to like each other more, compared to audio-only interactions. The importance of nonverbal behavior is also supported by the intuition of businesspeople who still conduct most important business meetings face-to-face rather than on the phone.

3.9.2 Relational Communication

Relative to the theory outlined above, conversation can be seen as an elaborate dance in which every utterance made by one partner contributes to his assessments of relational expectations and trust in the other, and thus to his perception of the relationship. According to Duck, every verbal message contains two elements: 1) propositional content and 2) a message
about the relationship; “You can barely utter a word without indicating how you feel about the
other” (Duck, 1998).

Every utterance must either fit squarely within the relational expectations that partners
have (serving to ratify the relationship by keeping current expectations high), or is to some
extent unexpected, in which case it should be marked and/or mitigated to avoid threatening
the face of one or both partners. Politeness theory provides some guidelines about the form
these marking and mitigating strategies should take (Brown & Levinson, 1987).

This is not to say that partners should use the exact same language in every situation, but
that at some level of abstraction the kinds of topics they discuss and the activities they engage
in with language has a regularity to it that is defined by their relationship. The exact form of
language used is, in fact, expected to vary subtly with the context (both immediate and
historical), and conversationalists who do not adapt their language in this way are seen as
inept and un-attentive to the conversation (Duck, 1991).

3.9.3 Social Deixis

One way in which language can be used to set relational expectations is through social
deixis, or what Svennevig calls “relational contextualization cues” (Svennevig, 1999), which
are “those aspects of language structure that encode the social identities of participants…or
the social relationship between them, or between one of them and persons and entitites
referred to” (Levinson, 1983). Politeness strategies fall under this general category (facework
strategies are partly a function of relationship), but there are many other language phenomena
which also fit, including honorifics and forms of address. Various types of relationship can be
grammaticalized differently in different languages, including whether the relationship is
between the speaker and hearer as referent, between the speaker and hearer when referring to
another person or entity, between the speaker and bystanders, or based on type of kinship
relation, clan membership, or relative rank (Levinson, 1983). One of the most cited examples
of this is the *tu/vous* distinction in French and other languages.

For English, Laver encoded the rules for forms of address and greeting and parting as a
(partial) function of the social relationship between the interlocutors, with titles ranging from
professional forms (“Dr. Smith”) to first names (“Joe”) and greetings ranging from a simple
“Hello” to the more formal “Good Morning”, etc (Laver, 1981).

Forms of language may not only reflect existent relational status, but may be used to
negotiate changes in the relationship, by simply using language forms that are congruent with
the desired relationship. As stated above, Lim observed that partners may change their
facework strategies in order to effect changes in the relationship (Lim). And, according to
Svennevig:

The language forms used are seen as reflecting a certain type of relationship between the
interlocutors. Cues may be used strategically so that they do not merely reflect, but actively
define or redefine the relationship. The positive politeness strategies may thus … contribute
to strengthening or developing the solidarity, familiarity and affective bonds between the
interactants. The focus is here shifted from maintaining the relational equilibrium toward
setting and changing the values on the distance parameter (Svennevig, 1999, pg. 46-47).
3.9.4 Social Dialogue

Common examples of social dialogue are small talk, conversational storytelling, gossip, getting acquainted talk, and joke-telling. What do these speech genres have in common? It is easiest at this point to define social dialogue by what it is not; it is talk whose focus is not on factuality and instrumentality, in short, talk that is not task-oriented. Given that all utterances carry relational meaning (as described above), what social dialogue is focused on is primarily the negotiation of the interpersonal relationship between the interlocutors.

Perhaps the purest form of social dialogue is what Malinowski referred to as "phatic communion", "a type of speech in which ties of union are created by a mere exchange of words." This is the language used in free, aimless social intercourse, which occurs when people are relaxing or when they are accompanying "some manual work by gossip quite unconnected with what they are doing" (Malinowski, 1923). An example of this type of small talk is the situation in which two strangers approach each other and the speaker (S) says to the hearer (H):

(1) Lovely day.

Utterances of this type have not received adequate treatment in linguistics or the philosophy of language. In these fields, "meanings" are taken to be those properties of utterances or texts which have their impact on hearers, such as describing states of the world or predicting or explaining human behavior (Devitt & Sterelny, 1999). The predominant, classical approach to meaning is that the meaning of a declarative statement is its truth-value when compared against the world. However, this captures very little of the meaning of utterance (1) in the situation described. The impact of this statement would be roughly the same if S made it under a very wide range of weather conditions, most of them far from ideal (although, if the weather is plainly very poor it could have an added element of irony associated with it). Thus, the classical notion of truth-value seems to play very little role in the meaning conveyed. In many situations, the impact of this statement on H would also be roughly the same if S approached H with a warm smile and gaze and uttered 'Lovely X', where 'X' could be composed of almost any short sequence of phonemes pronounceable by S. Thus, the meaning conveyed by this utterance in this kind of situation also has very little to do with its propositional content in the first place. The classical approaches to meaning do not help much in explaining the meaning of utterances in social dialogue.

As Wittgenstein and others have pointed out, there are many kinds of meaning. In addition to the classical, propositional notion of meaning described above, I claim there are two other kinds of meaning that are important for conversation: interactional and relational. Interactional meaning is that aspect of an utterance which affects what interlocutors do in the current interaction, affecting their behavior in such things as turn-taking, grounding, and emphasis (Duncan, 1974). Relational meaning is that aspect of an utterance which indexes the relationship between the interlocutors, and the rights and obligations entailed. The 'truth' of this meaning is the degree to which these obligations are adhered to (the degree of authenticity or sincerity).

Returning again to utterance (1), it can be seen to have: a propositional meaning (some congruence with the current weather conditions); an interactional meaning (transitions into a
conversation (Laver, 1975)); and a relational meaning (indexed a particular type of relationship) all at once. Thus, "small talk" can be seen as a conversational recipe for a particular kind of talk, and is thus fundamentally an interactional construct. One of its primary functions is to provide the opportunity for social dialogue, since it is a type of talk in which there is minimal commitment to factuality and minimal obligations entailed. However, its use does index a particular type of relationship; stereotypically one between strangers or acquaintances who are on polite (if not "friendly") terms with each other.

Researchers in linguistics, sociolinguistics and social psychology have hypothesized various functions that social dialogue might have. As mentioned above, Malinowski observed that some forms of social dialogue (i.e., "phatic communion") are used to "establish ties of union" or "to establish bonds of personal union between people brought together by the mere need of companionship" (Malinowski, 1923). Jacobson included a "phatic function" in his well-known conduit model of communication, which is focused on the regulation of the conduit itself (as opposed to the message, sender, or receiver) (Jacobson, 1960). Laver concluded that the underlying goals of phatic talk are to establish relationships and achieve transition (Laver, 1975). Coupland & Coupland hypothesized that phatic discourse involves a minimized commitment to factuality, open disclosure and seriousness, and that utterances could be classified along a continuum of "phaticity", rather than being discretely phatic or not (Coupland, Coupland, & Robinson, 1992). Dunbar found that the purpose of social dialogue is primarily to build rapport and trust among the interlocutors, provide time for them to "size each other up", establish an interactional style, and to allow them to establish their reputations (Dunbar, 1996).

In summary, social dialogue is conversation in which issues related to any instrumental task talk is backgrounded and interpersonal issues are (by default) foregrounded. Relative to the theory presented above, then, the role of social dialogue is primarily to build or maintain relational expectations and thereby trust in the relational partner.

### 3.9.5 A More General Model of Facework in Face-to-Face Conversation

In this section I present a dimensional model of relationship and a new model of facework that is used as the basis for a computational model of dialogue planning (presented in Chapter 4) and an evaluation of the effects of social dialogue on trust in relational agents (presented in Chapter 5). This work was previously published in (Cassell & Bickmore, 2001).

As discussed in section 3.1, there is an equivalence between dimensional and provisional models of relationship, in that dimensional models provide an abstract feature space within which specific provisional models can be placed. The dimensional model used in this theory of facework, while not providing the granularity of specific beliefs and intentions of the interlocutors with respect to their relationship, provides a good starting place for explorations in generating dialogue moves that are not only congruent with an existing relationship but serve to move the relationship in a desired direction.

#### Prior Models of Facework

As described in section 3.7.2, Brown and Levinson extended Goffman’s notion of face in their theory of politeness forms in language (Brown & Levinson, 1978). They defined positive face as an individual’s desire to be held in esteem by his interactants, and negative face as his desire for autonomy, and characterized the degree of face threat of a given speech
The face threat to the hearer can be given by the formula in Figure 3-1.

The ‘intrinsic threat’ parameter accounts for the fact that certain speech acts are more of a threat than others. For example, an informing is less of a threat than a request for information which is less of a threat than a rejection. Distance is defined to be a symmetric social dimension of similarity/difference within which the speaker and hearer stand for the purposes of this act and power is the ability of one interactant to control the behavior of the other.

If a significant threat will result from the speaker producing the indicated speech act, then the speaker has several options: 1) don't do the act; 2) do the act "off record"; 3) do the act "on record" with redressive action (negative politeness strategies); 4) do the act on record with redress action (positive politeness strategies); 5) do the act on record, "baldly". Following Grice’s (Grice, 1989) description of how to fail to fulfill the conversational maxims, these options are ranked in order of decreasing ability to mitigate a threat, thus the most threatening acts shouldn't be done at all, while the least threatening acts can be done baldly on record. Examples of "off record" acts are hinting and/or ensuring that the interpretation of the utterance is ambiguous (e.g., "I'm thirsty."). Negative politeness strategies include those which are oriented towards the autonomy concerns of the listener (e.g., "Could you bring me a drink?"), while positive politeness strategies address the esteem concerns of the listener (e.g., "Hey my friend, get me a drink.").

Svennevig extended Brown and Levinson’s model by noticing that the threat perceived from different types of speech acts can change based on context, and in particular based on the relationship between the speaker and hearer (Svennevig, 1999). For example, close friends have established a set of mutual rights and obligations and thus do not experience certain acts (such as requests) as face threatening, but rather as confirming and reestablishing their relational bonds. (This view actually fits the theory presented in this chapter very well.) Thus, his extension to the model can be characterized as shown in Figure 3-2.

![Figure 3-1. Brown & Levinson’s Face Threat](image)

![Figure 3-2. Svennevig’s Face Threat](image)
Svennevig’s dimensional model of interpersonal relationships has four dimensions:

- **Power** - the ability of one interactant to control the behavior of the other.
- **Solidarity** – degree of "like-mindedness" or having similar behavior dispositions (e.g., similar political membership, family, religions, profession, gender, etc.), and is very similar to the notion of social distance used by Brown and Levinson (Brown & Levinson, 1978).
- **Familiarity** - based on social penetration theory (Altman & Taylor, 1973), which claims to account for the establishment and growth of interpersonal relationships, this dimension describes the way in which relationships develop through the reciprocal exchange of information, beginning with relatively non-intimate topics and gradually progressing to more personal and private topics. The growth of a relationship can be represented in both the breadth (number of topics) and depth (public to private) of information disclosed.
- **Affect** - the degree of liking the interactants have for each other, and there is evidence that this is an independent relational attribute from the above three (Brown & Gilman, 1989).

**A New Model**

Based on analyses of interactions involving both task and social dialogue (initial meetings between real estate agents and buyers), I have extended Svennevig’s model of face threat to include measures of topic coherence and topic intimacy.

It is clear that the introduction of conversational topics that are at a significantly deeper level of familiarity than is expected relative to the existent relationship and activity will be seen as a face threat. For example, if a stranger on the street asked you how much money you had in your bank account, you would likely perceive this as a threat to your face. Such a kind of face threat is key to task encounters where strangers must interact, and occasionally share personal information. I term this a "Social Penetration" threat, or $\text{SP}_{\text{threat}}$.

Topics that are at the expected level of familiarity but which are completely unrelated to the topic at hand also seem to be face threats, but have not been accounted for in a general way in previous theory. While a subset of these have been addressed in Brown and Levinson's theory (e.g., rejections), moves which are deemed dispreferred based solely on their sequential placement in conversation cannot be accounted for, given Brown & Levinson's use of isolated speech acts as their point of departure. Instances of such "sequential placement" threats are failing to demonstrate the relevance of a conversational story, failure to appreciate conversational stories following their conclusion (Jefferson, 1978), or introducing conversational topics or stories which are not related to the on-going discourse (not "locally occasioned" (Sacks, 1995)). Thus, for example, if you tell someone a long humorous story, and they respond, not by nodding or otherwise acknowledging your story, but instead by changing the topic, that will threaten your face.

The resulting model of face threat is shown in Figure 3-3. It is novel in that it goes beyond the analysis of a single speech act to acknowledge and incorporate the role of discourse context into the determination of face threat. This model can be used by a relational agent to provide information about dialogue moves that it should or should not make in order to avoid threatening the face of its users.
3.9.6 Face-to-Face Conversation with Relational Agents

It should be clear that emulations of face-to-face conversation are the ideal medium for relational agents to use in order to build complex relationships with their users. The model of face threat above can be used on an utterance-by-utterance basis by a relational agent to inform its selection of next topics in this setting. Of course, many other factors need to be taken into account in selecting the content and form of a next utterance, including the attentional and intentional state of the dialogue (Grosz & Sidner, 1986), information structure (Prince, 1981), and the goals of the agent. Taken together, these factors allow an agent to meet the relational expectations of a user and thus build trust simply by engaging in dialogue with them. Chapter 4 presents a computational model that applies this model of facework to the problem of planning task and social dialogue in a particular type of service encounter, and Chapter 5 presents the results of an evaluation of the effectiveness of this model.

3.10 Application to the Macro-Structure of Long-Term Relationships

Relational partners cannot typically spend all of their time together. Given that actual enactment of relational provisions is the primary means of maintaining relational expectations, these expectations may decay over time when partners are apart, may be challenged by information from third parties, or may be subject to doubting brought about through cognitive re-appraisal processes. However, even with these potential threats to relationship longevity, most relationships seem to survive times when partners are apart, whether it be hours, days or even years. Something persists in their heads which lets them pick up where they left off the last time they were together. According to Duck:

\[
\text{Face threat} = f(\text{SA threat}, \text{SP threat}) \\
\text{SA threat} = f(\text{SA}_k, \{\text{SA}_1, \ldots, \text{SA}_j\}, \text{Power}, \text{Solidarity}, \text{Familiarity}, \text{Affect}) \\
\text{SP threat} = f(\text{FamiliarityDepth}, \text{TopicDepth}) \\
\]

Where,

\(\text{SA threat} = \text{Threat due to the speech act.}\)
\(\text{SP threat} = \text{Threat due to violation of social penetration theory.}\)
\(\text{SA}_k = \text{The class of speech act.}\)
\(\{\text{SA}_1, \ldots, \text{SA}_j\} = \text{The discourse context of speech acts into which SA}_k \text{ will be introduced. For example, SA}_1 \text{ could represent the overall conversation, and SA}_j \text{ represents the activity which SA}_k \text{ will become a constituent of.}\}
\(\text{TopicDepth} = \text{The “depth” of the topic to be introduced (wrt social penetration theory).}\)
\(\text{FamiliarityDepth} = \text{The current value of the depth dimension of the familiarity aspect of the relationship among the interlocutors.}\)

Figure 3-3. A New Model of Face Threat
We do not go through a ritual each breakfast time where we treat each other like strangers and run through a whole range of rewarding techniques to re-establish the relationship and take it to where it was the day before: we behave that mental way only with friends we have not seen for ages. The remarkable fact about daily life is that continuities exist in our minds and do not have to be worked for, once the relationship is defined and established (Duck, 1998).

How do partners maintain their relationships over the long term and especially when they are apart? What exactly do they need to remember about their prior interactions when they meet again? And, how do relationships normally change over time? Relational agents that are designed to interact with users for more than a single brief session—and especially those which live with their users for months or years—need to be designed with these issues in mind.

3.10.1 Relationship Maintenance

A good question phrased by Duck is, “Do relationships fall apart unless they are maintained, or do they stay together unless they are taken apart?” (Duck, 1988). While some relationship theorists weigh in on both sides of this question, it is clear that there are specific behaviors that partners engage in which serve to actively maintain their relationship.

Partners perform a wide range of behaviors in order to keep their relational expectations high. Following the theory presented above, the most fundamental way that partners do this is by simply performing the collaborative actions which constitute their relationship. This is supported by research finding that the most common maintenance behavior performed by couples is simply sharing tasks (Dainton & Stafford, 1993).

In addition to routine maintenance behaviors, many researchers have investigated “strategic maintenance behaviors,” which are those performed with the conscious intent of maintaining the relationship. Some of the most common categories of these behaviors include prosocial behaviors (e.g., telling a joke, initiating interaction), ceremonial and ritual behaviors, meta-relational communication (talking about the relationship), and assurances (explicit statements about commitment to the relationship) (Gilbertson, Dindia, & Allen, 1998) (Dainton & Stafford, 1993; Dindia, 1994; Stafford & Canary, 1991)

Routine maintenance behaviors, on the other hand, are those behaviors performed on a regular basis without the conscious intent of maintaining the relationship, but which serve this purpose nevertheless. Gilbertson, et al, investigated the use of “continuity behaviors” among cohabitating couples, which are behaviors enacted before, during and after an absence to bridge the gap in the relationship caused by the absence (Gilbertson et al., 1998). Prospective behaviors (enacted before an absence) includes such things as telling one’s partner what you will be doing during the time apart, affirmations, and farewells; introspective behaviors (enacted during an absence) includes communicating with one’s partner (e.g., via telephone), and talking about them to others; and retrospective behaviors (enacted after an absence) included greeting and asking what your partner did during the absence.

3.10.2 Expected Trajectories of Relational Change

In addition to simply maintaining relationships, people generally expect relationships to change over time, in predictable patterns. According to Duck (Duck, 1988) as a relationship develops between individuals, the following things change:
• The content of interactions and the activities the partners engage in change.
• The diversity of interactions (the number of different things partners do together) generally increases.
• The qualities of interactions change, such as language forms and nonverbal behavior in face-to-face interaction.
• The relative frequency of interactions increases.
• Reciprocity decreases and complementarity increases. Reciprocity is when partners respond to each other’s actions in similar ways (e.g., feeling obliged to repay social goods in kind); on the other hand, complementarity is when partners respond in different ways that fit together (e.g., one dominates and the other submits).
• Intimacy (physical and psychological) increases.
• Interpersonal perception aligns with self-perception. Partners’ perceptions of themselves tend to align with their partner’s perceptions of them.
• Commitment increases. Commitment is intention to continue in a relationship, and is a general measure of the strength of a relationship.

Several researchers have proposed “stage models” of relationship, which assume there are a fixed set of stages that different types of relationships go through. Examples of stage models are those by Lewis (Lewis, 1972), who proposes that (marital) relationships begin with similarity-based attraction then proceed through stages of good rapport, mutual self-disclosure, empathic understanding of the other person, role compatibility, and finally commitment to the relationship and identity as a couple. In another model, Reiss proposes that relationships go through four stages: initial rapport; mutual self-revelation; mutual dependency; and personal need fulfillment (Reiss, 1960). Stage models are now generally considered to provide very weak predictive power given their assumption of a fixed sequence of stages, since actual relationships often jump around among various stages in a non-linear manner (Brehm, 1992).

3.10.3 The Role of Personal History

According to the theory above, knowledge of past times together is only important for informing assessments of expectations of future accommodation behaviors; once these assessments are complete this knowledge can be discarded. Yet memory of specific past interactions seems to be very important to relational partners. For example, Planalp found that mutual knowledge and talking about past events were among the primary features that observers used to reliably differentiate between conversations involving friends and those involving acquaintances (Planalp, 1993). Exactly what role do these memories play, and do they represent an important feature to build in to relational agents who will be interacting with their users over long time spans?

As mentioned above, relational expectations can decay, especially in the absence of one’s partner. Thus, one important role of remembering, and talking about, past activities together, is to keep relational expectations high. For example Stafford and Canary found that one of the maintenance strategies couples use is to remind each other about relationship decisions made in the past (Stafford & Canary, 1991).
As relationships develop, it is clear that partners gain knowledge about each other, both through informing and sharing experiences; increasing “common ground”.

This is not to say that partners’ memories of past shared experiences are perfect, or that they even agree on them. Even events as important as first meetings and the ways in which a relationship started (“origin stories”) are often remembered very differently by partners early in the relationship (Duck, 1991). Miell, in a study of weekly assessments of relationship strength over several months, found evidence to support the view that “memory for relationships … as being an active, interpretive process, where the events of the past are reconstructed, rather than reproduced, in the context of current events” (Miell, 1987).

In conclusion, memory of one’s relational partner and specific past activities together is an important part of relationships, in that this memory provides a “common ground” or “shared meaning system” which lets dyads communicate more efficiently, re-evaluate their relational expectations, and provides fodder for certain maintenance activities such as reminding each other of past shared activities or relational decisions. Retention of this knowledge could be viewed as a type of accommodation behavior; partners routinely expect each other to interpret their utterances in part by filling in what is presumed to be shared knowledge. However, I still contend that relational memory is not an intrinsic part of the concept of relationship; you can know a lot about someone without having a relationship with them and you can have a relationship with someone (e.g., contractual) without any relational memory of past interactions.

### 3.10.4 Persistent Relational Agents

In order to build relational agents capable of multiple interactions with users over long periods of time it is clear that they should have at least some of the capabilities outlined above, including: some memory of information about the user and past interactions; some changes in relational behavior over time in ways expected by the user; and the use of routine and strategic relationship maintenance behaviors. Chapters 7-9 present the design and evaluation of a relational agent with these capabilities.

### 3.11 Summary

In this chapter I have outlined a theory of what a relationship is, some of the strategies that relational agents should use in making relational decisions, and a range of relational behaviors that can be used by them to build, maintain and evolve relationships with their users. I also discussed how this theory could be applied to the design and evaluation of relational agents in two realms: the micro structure of face-to-face conversation and the macro structure of long-term relationships. In the following chapters I will present implementations and evaluations of relational agents in each of these areas.

---

1 Clark defines the “common ground” of two interlocutors as the sum of their mutual knowledge, mutual beliefs, and mutual suppositions (Clark, 1992). My use of this term is slightly different. Here, I am interested in the mutual knowledge that relational partners persistently carry over time, particularly between interactions.