A COMPUTATIONAL MODEL FOR BUILDING RELATIONSHIPS BETWEEN HUMANS AND VIRTUAL AGENTS

by

William Coon

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Master of Science

in

Computer Science

October 2012

APPROVED:

Dr. Charles Rich, Major Advisor ______________________________

Dr. Candace Sidner, Major Advisor ______________________________

Dr. Sonia Chernova, Reader ______________________________

Dr. Craig Wills, Head of Department ______________________________

Abstract

As artificially intelligent agents become more advanced, they will require corresponding advances in social capability. In particular, they will require an understanding of the development of relationships. This work is intended to aid in addressing this need. We have developed a model of the development of relationships, designed and implemented a planning module based on this model, and performed an evaluation study verifying the functionality of the model and implementation. This should provide a foundation for future work in developing artificially intelligent agents capable of appropriately dealing with the development of social relationships.

Contents:

Acknowledgements

I would like to sincerely thank my advisors, Professors Charles Rich and Candace Sidner, for their guidance on this project and my work as a student and a member of the Interaction Lab. They have given me invaluable support, advice, and direction.

I would also like to thank my reader, Professor Sonia Chernova, for her comments and suggestions on this project. She provided an outside view on the project that helped me clarify and express these results.

My gratitude also to Professor Joseph Beck for his advice on the evaluation study portion of this thesis, and his help in selecting appropriate metrics and methods of analysis.

The students of the Interaction Lab have also been crucial in supporting me as I explored this subject and developed this thesis. I thank them for the advice and motivation they provided.

Finally, I would like to thank the students and faculty of Worcester Polytechnic Institute, my second home for the duration of this work.

This work is supported in part by the National Science Foundation under awards IIS-1012083. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Table of Figures

1. Introduction

The ubiquitous human tendency to anthropomorphize is a constant in the field of artificial intelligence. Humans develop certain expectations of how social interactions will work, and they apply these expectations when interacting with complex systems, particularly artificially intelligent systems. In particular, humans expect advanced artificially intelligent agents to have an understanding of relationships and be able to reason about them. In humans, this is a fundamental skill: learning how to alter behavior to suit a particular relational context. We have developed relationship models and algorithms employing them in order to provide agents with this capability, and thus to work alongside humans for extended periods of time in a more natural fashion. The primary component of our work is a model of relationship stages and progression. This component provides information on several levels (informing degree of familiarity and mode of address, for example), but the fundamental use of the model is in planning of conversations and activities. Therefore, the relationship model is accompanied by a planning model and algorithm, designed to select activities appropriate to a given relationship while also attempting to accomplish goals and advance the relationship. This work will enable the construction of virtual agents and robots that can develop relationships with humans over time.

This research was conducted in the context of the Always-On Relational Agent project (Nooraei, Liu, Conley, Sidner, Rich, and Bickmore 2011). The goal of this project is to create a virtual agent companion capable of assisting with various tasks and functioning in a social context while developing a relationship with a human. The current focus of the project is to develop a companion-assistant for isolated older adults. This agent is designed to engage in social activities, such as card games, help navigate social media, such as Skype™, manage a schedule, and encourage exercise and the development of other social relationships.

1.1 Related Work

In developing the work reported in this thesis, we drew upon sociological and psychological research for qualitative models and expositions of the phenomena we are attempting to formalize, as well as artificial intelligence research that addressed social and relational issues.

The sociological and psychological literature is clear on the point that humans not only exist in a social context, but also think and process information in ways that depend strongly on social factors. In particular, social cognitive neuroscience is a discipline concerned with the interaction between social cognition and other levels of thought. Hardin and Conley (2001) provide a useful introduction to this subject, describing how epistemic thought is grounded in a social environment and contrasting the social cognition approach with earlier, more naïve approaches. They assert that humans judge the truth of beliefs largely by comparison to the beliefs of others, and that relationships require a common “shared reality” (or set of grounding beliefs about the world). Ochsner and Lieberman (2001) provide a good overview of social cognitive neuroscience, including descriptions of the different contributing disciplines and the major realms of inquiry, as well as some significant results (regarding stereotypes, attitude changes, social perception and inference, self-knowledge, and the relationship between emotion and cognition). Nickerson (2001) describes the phenomenon of “projective knowing,” in which others’ knowledge is modeled as a series of additions to and subtractions from a base of one’s own knowledge. Naturally, this method, though common, leads to a number of problematic assumptions that can cause significant misunderstandings.

Our methodology involves studying a phenomenon within human interaction, producing a computational model, and implementing it in an agent. Other artificial intelligence researchers are having success with similar methodologies, implementing models based on various sociological and psychological phenomena, such as the establishment of rapport in communications (Gratch, Okhmatovskaia, et al. 2006). Gratch and Okhmatovskaia demonstrate that rapport-establishing techniques based on human interaction can be used effectively with virtual agents, and that there are practical benefits to doing so. Rich, Ponsler, Holroyd and Sidner (2010) studied the phenomenon of engagement, the process of starting, ending, and managing the perceived connection between conversational partners. They developed computational models of this process, then implemented general-purpose engagement modules, as well as a full implementation in a game-playing robot.

The model at the core of this paper will serve as the basis for the long-term reasoning portion of the Always-On Relational Agent. It will be responsible for tracking the development of the relationship over the course of weeks and months as well as planning daily activities and lines of conversation. The real-time execution of these plans, and other more direct aspects of interaction, will be handled by the Collaboration Manager, a separate component being developed in parallel by Bahador Nooraei. The architecture of this system is more thoroughly described in Section 4.1.

There are currently several other projects in various stages of development that are similar in focus to the Always-On Relational Agent. For example, Shibata, Kawaguchi, and Wada (2012) studied the interactions of people with PARO, a plush seal robot, over extended periods of time. They collected data on the activities people engaged in, as well as their impressions of the robot, through the course of their ownership. However, their robot was unable to engage in conversation, and the aim of their study was to collect demographic data related to pet robots, rather than to focus specifically on the mechanics of social interactions. Project Sera (Klamer & BenAllouch, 2010) was an extended investigation, now completed, of various aspects of social interaction with artificially intelligent agents, similar in several ways to the Always-On Relational Agent. In particular, some of their studies specifically focused on social relationships between older adults and robots staying in their home for extended periods. Although these robots did not generally have planning capabilities, and were single-function, these studies clearly demonstrated the development of attachments and relationships between humans and artificial agents. Turunen et al. (2011) have also investigated and developed live-in agents that serve as conversational partners, focusing largely on dialogue construction and planning. The focus of this study differs from the Always-On Agent, in that these agents are largely specialized in a particular domain (health and exercise, in this case), and were not equipped with general-purpose relationship-building capabilities.

1.2 Overview

In Chapter 2, we will provide an overview of the literature, primarily in the fields of sociology and psychology, that has influenced or guided the development of our mathematical models of relationship. In Chapter 3, we will present the mathematical models themselves, in abstract form, which underlie the development of the relational virtual agent. In Chapter 4, we will present the implemented form of the models. In Chapter 5, we will walk through an example run session of the agent's planner, in which a plan is constructed, executed, and revised. In Chapter 6, we will present an evaluation study we performed to test the plausibility of the models and the feasibility of the planner. In Chapter 7, we will discuss our conclusions and possible future directions for related research. Finally, the appendices provide the materials used in our study as well as the sample activity metadata used in the study and the example session.

2. Psychological and Sociological Theories of Relationship

The literature in psychology and sociology has provided a basic foundation for the computational model of relationships developed in this work. The model is a synthesis of our understanding of this literature. In particular, one of our major challenges was to develop quantitative structures to capture the primarily qualitative observations of this literature.

Firstly, various authors in this literature provide a general basis for understanding the underlying social territory our agent is to be expected to navigate. Cole and Bradac (1996) describe many factors perceived as contributing to successful relationships. They conclude that relationship satisfaction primarily relies on emotional closeness and intimacy, as well as knowledge of the boundaries of emotional intimacy for the relationship. Other significant factors include the relative degrees of activity or spontaneity and the demeanors of the people in the relationship.

Wish, Deutsch, and Kaplan (1976) explore the various dimensions along which relationships can vary. Their findings cover a wide variety of relationships and relationship qualities. They note that close friendships tend to be characterized by a cooperative, friendly atmosphere and a general sense of equality, but also by an intense emotional closeness. Moreover, close friendships tend to be informal and focused on social and emotional matters, rather than concrete tasks, which require a more formal relational approach. Naturally, there is a fair amount of variance on these dimensions from person to person, as different people are better suited to different relationship styles.

Burgoon and Hale (1984) provide an overview of the various topics involved in relational communication. They examine literature relevant to several different perspectives on this issue, synthesizing findings in different areas, including primitive expressions of dominance, the semantic content of relational behaviors, and interpersonal evaluations. Moreover, they break down the ways in which relational communication relates to the development of relationships, such as signaling the desire to increase the level of closeness, or developing “smoothness” of interaction.

Stafford, Dainton, and Haas (2000) catalogue behaviors involved in relational maintenance, including routine behaviors and those that are planned specifically for relational maintenance purposes. These include explicit assurances about the significance of the relationship, bids for intimate conversation or advice, conflict management, collaborative tasks, general positivity, and social activity with common associates. Many of these behaviors are part of the design of the Always-On Relational Agent. Gilbertson, Dindia, and Allen (1998) also studied maintenance behaviors, with a focus on the establishment and maintenance of continuity in a relationship during periods of separation (of varying duration). They conclude that prospective behaviors (those preceding and mitigating the separation, such as informing the other person how long you expect to be gone, planning activities for afterward, or simply saying “good-bye”) are significantly positively correlated with relational satisfaction.

Beyond these explorations of the general underpinnings to the model developed here, Social Penetration Theory (Altman & Taylor 1973) has been a significant influence on the development of the particulars of our model. Under this theory, the development of a relationship is described in terms of the increasing breadth and depth of topics available for discussion: as two people grow closer, they are able to discuss more intimate topics, as well as a broader range of topics of a given intimacy. Upon first meeting, topics are mostly limited to “small-talk,” e.g. discussions about the weather, sporting events, nearby occurrences, or local news. Later on, after becoming more familiar, and in fact as part of the process of becoming more familiar, people begin to reveal (generally non-radical) personal opinions, such as their general political alignment. Eventually, the range of acceptable topics comes to include more personal matters, such as relationship troubles or extreme viewpoints.

We have drawn also on the artificial intelligence work of Bickmore (2003), extending his notion of "accommodation" to the relationship framework. Bickmore's idea is that people who collaborate develop a mutual expectation that certain tasks will come to be accommodated, meaning that whenever appropriate circumstances arise, each person will provide the relevant assistance without having to be explicitly requested. For instance, if two people are assembling tables together, they will learn the required steps and hand over tools when needed, or one may hold a part of the table steady while the other is working. A natural extension of this idea forms a key component of our model, i.e. a set of "available" activities that grows as the relationship develops.

Bickmore and Picard (2005) present a differing relationship model based around three dimensions: depth of self-disclosure (equivalent to the concept of depth within social penetration theory), breadth of information known about each other, and solidarity (the degree to which the two people have similar behavioral dispositions). Our model differs from this one, as it was developed in a quite different context. The Bickmore/Picard model was intended for use in a real-estate agent, focusing on a sales-related task and utilizing reasoning about relationships to perform this task more effectively and naturally. Our model is intended to be of more general design, having relationship development as an intrinsic rather than a utilitarian goal, and remaining highly modular with respect to capabilities and non-relational values. That is, the relationship model is designed to function in virtual agents and robots of varying design, independent of a specific goal system or set of capabilities.

Ultimately, as will be described in the following chapter, the model developed here is designed to address many of these concerns, with a particular focus on the topics and activities available at different points in a relationship, as in the social penetration model, using a method that draws significant inspiration from the accommodation model. Of course, this model is only a starting point. We have had to focus on those areas of relationship research most germane to our goals, and as such there is a great deal of territory left for future research to cover.

3. Computational Model

The system I have developed rests on a number of interrelated computational models of different aspects of social interactions. At the core is the model of relationship stages, archetypal relationships that serve as reference points both for the agent and for the human interacting with it. Around this structure, we connect relationship progression with planning of activities using a model of closeness and utility, augmented by a model for avoiding repetition and a model of conceptual relevance.

3.1 Relationship Stages

We use relationship "stages" to represent points along the progression of a relationship, markers of the relationship's depth. They serve as archetypes, with attendant permissions and requirements that humans can be expected to intuitively understand. For instance, two people might be "strangers" or "acquaintances". This system of classification provides information to a social agent about how to interact with people (e.g., the expected level of formality) as well as what sorts of activities are appropriate to engage in.

This notion of stages is built partly on the notion of "accommodation" (Bickmore 2003) discussed in Chapter 2. Based on that framework, we could consider relationship stages as formally equivalent to the set of activities permitted within a given relationship, given some particular universe of possible activities. The containment order on the permitted activity sets would then generate a natural lattice of stages. Generally, the relationships with larger permitted activity sets are more intimate than those with smaller sets, and the progression from smaller to larger sets of activities is accomplished by establishing greater closeness. See Figure 1 for an example relationship lattice over the activity universe {Singing, Dancing, Philosophizing}. This universe permits eight different relationships, one for each subset of the activity set. The most "distant" relationship would naturally be the "empty relationship", at the bottom of the lattice, wherein no activities are permitted. On the other end of the spectrum, and the top of the lattice, is the relationship in which *all* activities are permitted. Other relationships lie between these extremes.

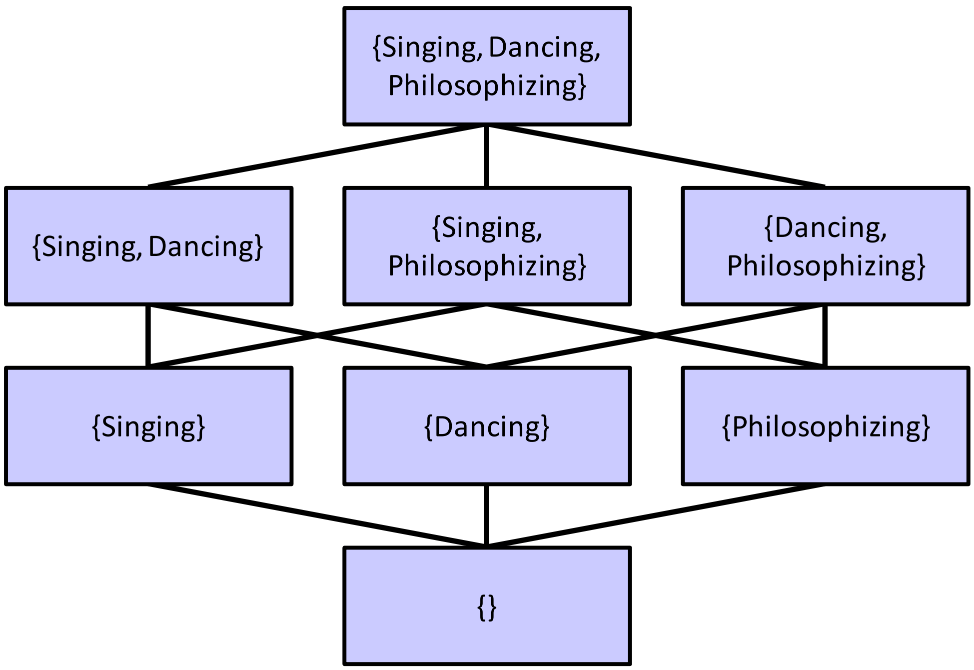

Figure 1 – Basic Relationship Lattice

Tracking the level of closeness as well as the permitted activities allows the relationship stage model to both "push" and "pull". On one hand, when evidence arises that a relationship has become closer, this may indicate that the field of socially appropriate activities has widened, e.g., if you spend a great deal of time interacting pleasantly with a coworker, you may decide to spend time with them outside of work. On the other, if a bid to expand this list is successful, this may indicate that the relationship has progressed, e.g., if someone accepts your request for a date, the expectations for your relationship shift rapidly.

Under normal conditions, however, a full lattice as in Figure 1 would be much too fine-grained. In practice, we select specific points in this relationship lattice, based on common relationship patterns. This makes the relationship progression more natural to those interacting with the system, and facilitates authoring decisions, as we will describe later in this section. In the general case, the progression remains a partial order – that is, there may be pairs of relationships with permitted activity sets that are neither subsets nor supersets of each other. For a small example, see Figure 2. As illustrated, it is common to discuss extremely personal topics with one's doctor, but it would be unusual to invite them to see a film with a group of friends, whereas this situation is reversed with some acquaintances. As we will discuss in Chapter 4, the relationship lattice for the Always-On Relational Agent is a straightforward one, consisting of the relationships Strangers, Acquaintances, and Friends, in a linear progression (see Figure 7).

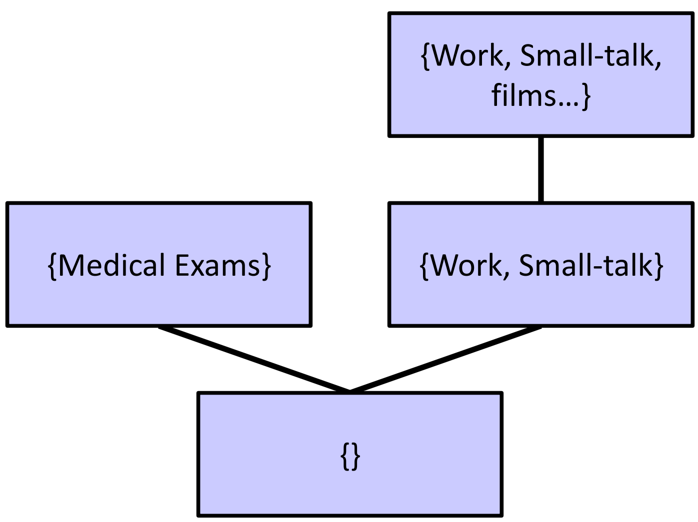

Figure 2 – Pared-Down Relationship Lattice

Thus, as the virtual agent and the human interact and expand the scope of their interactions, eventually the agent will reason that their relationship has progressed to a new stage, which brings with it new permitted activities, modes of address, and other expectations.

In most practical cases, of course, it will not be possible to exhaustively list the activities that characterize a given relationship stage. The Always-On Relational Agent in particular is designed as a general framework to which new activities are incrementally added via activity plugins. As such, in these cases relationships can instead be characterized by a set of guidelines for activity plugin authors, leaving the categorization of activities to be performed by those with domain-specific knowledge. Thus, when specified each relationship stage is associated with a qualitative description rather than an exhaustive activity list, but individual activities are still ultimately associated with stages by plugin authors.

In this case, the conceptual structure of the model remains unchanged: relationships still may be thought of as points in a lattice of subsets of all possible activities, even though the activities may not be specifically listed ahead of time. Moreover, the ability of the model to both "push" and "pull" is not lost. Just as the relationship stage may still gate access to particular activities, so may the acceptance of activities categorized as "unavailable" (or the initiation of such activities by the human) still be used as evidence of a relationship shift. This can be accomplished in a straightforward way by increasing the value corresponding to the progression from the current relationship stage towards stages that permit the activity in question. In a linear lattice, this manifests simply as more rapid progression along the only available path, while in a more complex lattice this will increase the likelihood (as well as rapidity) of progression along all paths leading to relationships permitting the activity. In inverse cases, wherein activities corresponding to the current stage are repeatedly *rejected*, the progress of the relationship can be reduced, potentially resulting in a return to an earlier stage. To prevent the model from backsliding to a previous stage immediately after progressing to a new one, the level at which the backwards transition occurs is below that at which the corresponding forward transition occurred.

3.2 Closeness and Utility

One of the chief duties of the relationship module is to produce plans for the interactions the agent will have, and recommend which activities to engage in. To this end, we have developed a model of the major considerations it will have to take into account.

Generally, we divide activities into three groups: those that are *inappropriate* to the relationship at its current stage; those that are *available* to the relationship but not accessible at the current moment; and those that are currently *accessible*. The distinction between activities that are appropriate or inappropriate to the relationship is based on the relationship stage, as described in Section 3.1. To decide whether an activity is appropriate at the current moment, we define a level of required “comfort” or “temporary closeness”. Some activities require a greater degree of care, lead-in, and / or familiarity, such as bringing up sensitive topics of conversation. This closeness value is generally at a baseline level at the beginning of a conversation (unless another conversation has occurred very recently) and increases through social activity over the course of the interaction. The baseline temporary closeness level is determined by the overall stage of the relationship. Thus as the relationship advances to closer stages, more sensitive subjects can be broached more easily.

Once it is determined what activities are possible at a given moment, the question remains as to which is *best* to perform. We solve this with an expected utility calculation based on several concerns. The relational agent serves multiple purposes: in addition to being a social companion, it also provides more concrete benefits: it can help a person use certain software, organize a schedule, or play games. As such, we divide utility into two sources which we call "virtuous" and "social" utility, where the total utility of an activity is the sum of its virtuous and social utility values, as shown in equation 1. In this equation, S_D and V_D are damping constants, as described in Section 3.4.

$$\boldsymbol{U(A) = S_D S(A) + V_D V(A)} \qquad \text{(Eq. 1)}$$

Social utility (*S(A)*) is a measure of the social benefit of an activity. This quantity serves multiple purposes: it directly contributes to the utility of an action (as social interaction is a valuable feature of the agent), and it adds to the level of temporary closeness. This value is increased from the baseline by engaging in social activities, then gradually degrades back to the baseline level over time (see equation 2, described below, for the dynamics of this process). Virtuous utility (*V(A)*) refers to non-social value, any benefit derived from accomplishing goals of the agent that are not social in nature. Depending on the specific agent, these goals might include helping a person with organization or weight-loss. In practice, most activities have a mixture of both kinds of utility, though some are more heavily weighted towards one kind or another.

For example, helping someone with their tax returns would provide a large benefit, but would be minimally social, and so it would have a high virtuous utility and a low social utility.

Idle discussion on inconsequential topics, such as the weather, provides little or no benefit, and a fairly low level of social interaction, so it would have very low virtuous utility, and low-to-moderate social utility. Discussing personal decisions about healthcare regarding a serious illness might provide high levels of both social and virtuous utility, as it is very beneficial to make good decisions in this domain, and discussing such personal topics requires and strengthens intimacy.

For further examples of the values assigned to activities, see Appendix B. This appendix lists activities used in the example planning scenario (Chapter 5) as well as the evaluation study (Chapter 6).

Equation 2 provides the general formula for determining the temporary closeness at a time t_i, where $C(t_i)$ is the temporary closeness, S_A is the social utility provided by the activity A most recently performed (at time t_{i-1}, with $C(t_{i-1})$ being the temporary closeness prior to performing A), C_0 is the baseline temporary closeness, and k is a damping factor. What this means is that performing an activity raises the temporary closeness, but this value degrades to the baseline value asymptotically if no activities occur for an extended period of time. Thus, normally an interaction will begin at the baseline level of temporary closeness for the current relationship stage, but, when some level of temporary closeness is built up, it can still be used if follow-up interactions occur soon after.

$$\boldsymbol{C(t_i) = \frac{C(t_{i-1}) + S_A - C_0}{(t_i - t_{i-1} + 1)^k} + C_0} \qquad \text{(Eq. 2)}$$

Figure 3 shows an example of the decay curve of temporary closeness. In this example, the value starts at 100, the baseline value is 30, and the decay constant is 1.

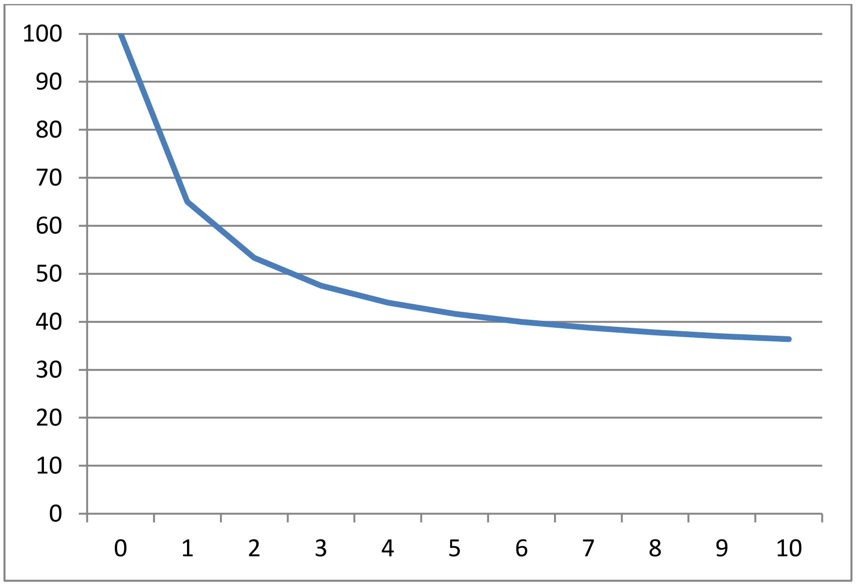

Figure 3 – Example Temporary Closeness Decay Curve

3.3 Performing Unavailable Activities

Until now we have been assuming that a social agent remains within the boundaries of an established relationship stage when making plans for interactions. However, with humans, a change in a relationship is often precipitated by an explicit bid to engage in an activity that was previously outside the relationship's bounds (Lim 1994), e.g., asking someone out on a date. To represent this, we permit the agent to plan such activities when sufficient mitigating factors are in play, including overall closeness, "temporary closeness", and the presence of activities in the permitted set that are associated with the same relationship stage as the "unavailable" activity.

3.4 Utility Damping

It is usually unnatural to perform the same activity repeatedly, as activities often intrinsically exhibit diminishing returns, or participants become bored. Moreover, if an agent proposes an activity and is rebuffed, it should wait a while before proposing the same activity again unless it is of very high priority. Thus, we apply damping factors to the utility of actions recently proposed or performed. We do this by reducing the value of specific activities by damping coefficients between zero and one that over time return gradually to a value of one, based on the recent history of the activity. This damping functions only on a short-term scale. Longer-term shifts in utility must be handled by system components with domain-specific knowledge, not the general planner. For example, if a request to play backgammon is consistently turned down, the agent should learn that the human probably does not enjoy this game, and should therefore permanently (modulo additional learning) lower the utility values of the activity. If, on the other hand, an offer of food is repeatedly turned down, the agent should *increase* the expected utility of eating.

This damping principle is also applied to the two categories of utility on a wider scale – that is, particularly in the case of repeated rebuffs, it may be appropriate to (temporarily) reduce the valuation of either social or virtuous utility in proportion to the respective utility values of the rejected activities. For instance, if social topics are repeatedly rejected, then it may be wise to focus on more neutral activities for the time being.

Equation 3 shows the idealized general form of the activity-specific damping function. In practical implementations, coefficients sufficiently close to 1 or those derived from occurrences sufficiently far in the past may be discarded. In this equation, we assume that there have been *n* prior instances of the activity under consideration, indexed by *i*. *$U_D(A)$* is the dampened utility of the activity *A*, *U(A)* is the *a priori* utility, *k* is a decay constant, d_1 through d_n are the durations elapsed since previous instances of the activity (with subscripts corresponding to the indexing of the instances), and C_{ACCEPT} and C_{REJECT} are separate damping constants for accepted and rejected instances.

In intuitive terms, what this equation does is provide a coefficient less than 1 for each instance of the activity, possibly differing for accepted and rejected instances. The utility of the activity is penalized by each such coefficient. Over time, each coefficient approaches 1 at a rate determined by the decay factor, as shown in the equation. That is, the utility of an activity is multiplied by a coefficient less than 1 for each time in the past that the activity was performed or rejected, the precise value of a given coefficient being given by the time since the activity attempt occurred and whether it was successful. In a practical implementation, of course, we may adopt the simplification of discarding coefficients that come sufficiently close to 1.

$$\boldsymbol{U_D(A) = U(A)\prod_{i=1}^{n}\left(1 - \frac{C_{f(i)}}{(d_i + 1)^k}\right)} \qquad \text{(Eq. 3)}$$

$$\boldsymbol{f(i) = \{ \begin{matrix} C_{ACCEPT}, if\ instance\ i\ was\ successful \\ C_{REJECT}, if\ instance\ i\ was\ rejected \end{matrix}} \qquad \text{(Eq. 4)}$$

The social- and virtuous-utility damping formulas (equations 4 and 5) are similar to this activity-specific case. Each provides a damping coefficient to be applied to all calculations of social or virtuous utility, respectively. The chief difference is that each sub-coefficient is scaled by *s(i)* or *v(i)*, the utility of the appropriate type for the activity indexed by *i*.

$$\boldsymbol{S_D} = \prod_{i=1}^{n}\left(\mathbf{1} - \frac{\boldsymbol{C_{f(i)}s(i)}}{(\boldsymbol{d_i} + \mathbf{1})^{\boldsymbol{k}}}\right) \quad \text{(Eq. 5)}$$

$$\boldsymbol{V_D} = \prod_{i=1}^{n}\left(\mathbf{1} - \frac{\boldsymbol{C_{f(i)}v(i)}}{(\boldsymbol{d_i} + \mathbf{1})^{\boldsymbol{k}}}\right) \quad \text{(Eq. 6)}$$

Thus, these coefficients scale the two components of the calculated utility of an activity, as given by Eq. 1:

$$\boldsymbol{U(A)} = \boldsymbol{S_D S(A)} + \boldsymbol{V_D V(A)} \quad \text{(Eq. 1)}$$

3.5 Relevance

As an additional guiding principle for the planner, to create plans that "flow" more naturally, we model the relevance of topics throughout the session. Conversations should move from topic to topic in a reasonable manner, rather than randomly jumping about. As we will describe in Section 4.3.2, plans will consist of multiple possible paths of high utility. We wish to leave options for the agent and the human they are interacting with in determining the direction of the interaction. Therefore, topic relevance does not modify utility calculations, but instead acts to prioritize amongst high-utility plans.

Necessary for this judgment is a notion of which topics are relevant to the agent under consideration, and how they relate to each other. For this, we use a mathematical graph of these topics, general and specific, where the topics themselves form nodes, and topics linked by edges are directly related to each other as conversational topics. For an example, see Figure , which gives a partial topical relevance graph for an agent concerned with films. Nodes include specific films, genres, writers, directors, and actors. Edges link relevant topics in a natural way – actors with films they starred in, writers with films they wrote, etc. In this graph, we can see that actors Stan Laurel and Oliver Hardy are relevant to the film *Big Business* – this is because they had starring roles in this film, though the specifics of the association are not noted in the graph.

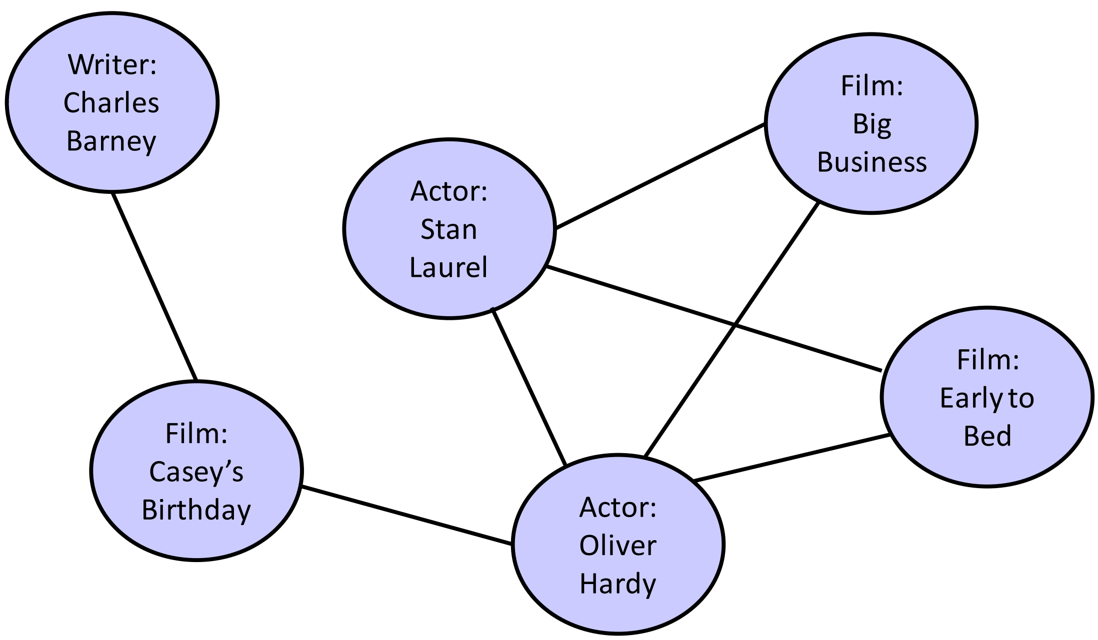

Figure 4 – Example Relevance Graph

The purpose of this graph is to prioritize *activities* in particular contexts, based on a preceding activity. Topics are distinct from activities, serving as general subjects activities might relate to. An activity might be associated with several topics, and several activities might all be associated with the same topic. For example, in the film-discussion agent, an activity might be "discuss the acting techniques in the film *Early to Bed*", which would be associated with the topics of "*Early to Bed*," "Stan Laurel," and "Oliver Hardy." The association of activities with topics is performed by someone with domain-specific knowledge.

The general form of the prioritization algorithm is that activities activate nodes they are associated with, and that this activation spreads (with reduced effect) to adjacent nodes. Therefore, given an activity, we have a level of activation for each node of the graph, corresponding to the relevance of each topic at the point in the plan after this activity is completed. Then, we can prioritize activities that are associated with more strongly activated nodes. A natural simplification of this model is to constrain spreading to only adjacent nodes, and to treat the activation level as a binary instead of a scalar value – that is, an activity is relevant if it is associated with a node or nodes identical with or adjacent to those associated with the prior activity. This modification of the model allows a sharp delineation between topics that are related and those that are not – and therefore between *activities* that are related and those that are not. Given this information, it is a simple matter to prioritize relevant activities over those that are irrelevant.

In more precise terms, the relevance algorithm works as follows. Let $G = (V, E)$ be the graph of topics, where V is the set of vertices and E is the set of edges. Let T be a function (the topic association function) from activities to subsets of V. Let A_0 be the activity just completed, and let C be the set of activities that are candidates to occur next. Then, the sets R and I of indices for activities that are relevant and irrelevant, respectively, are defined as follows:

$$R = \{A \in C \mid (\exists (a, b) \in E) \wedge \left(A \in T(A_0)\right) \wedge (b \in T(A)\}$$

$$I = C - R$$

For example, in Figure we see the state of the graph after discussing the actor Stan Laurel: the topic associated with this activity is shown bolded and encircled, and adjacent topics are shown bolded and connected by higher-weight, dashed lines. Discussions related to Stan Laurel, Oliver Hardy, Big Business, or Early to Bed would all be relevant.

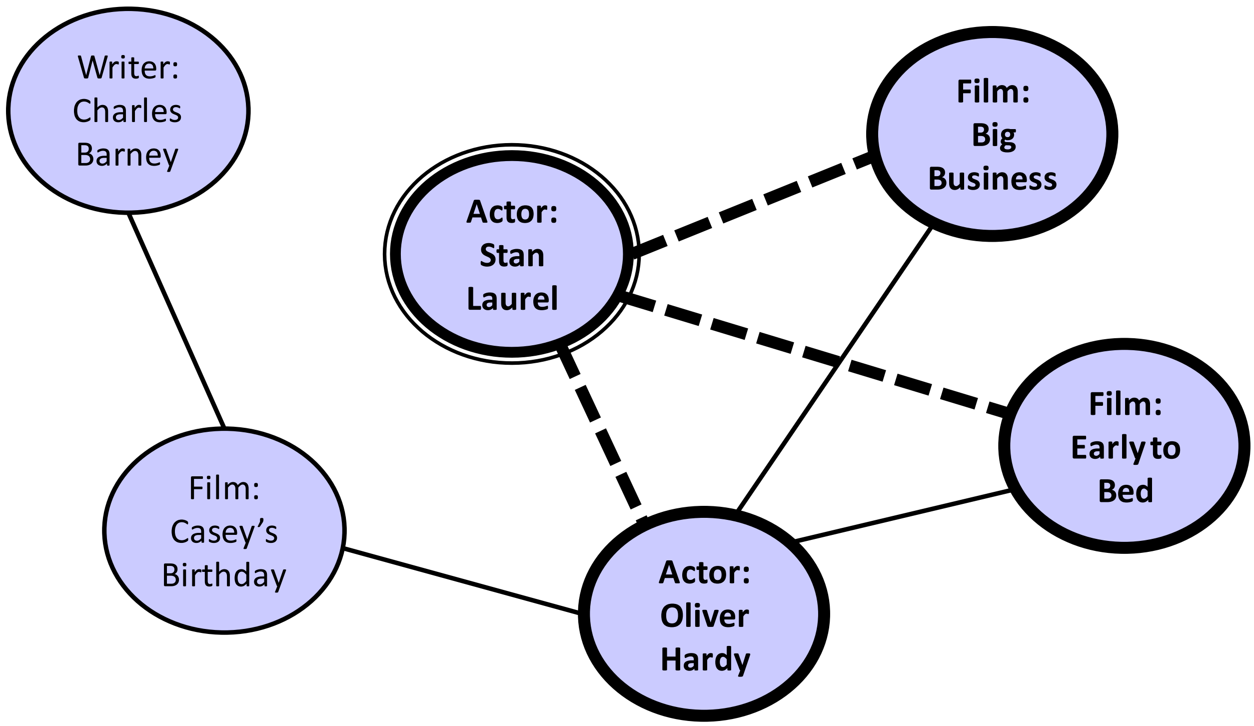

Figure 5 – Relevance Graph Activation Example

4. System Design

The models discussed in the preceding chapter are instantiated in an implementation of the Relationship Manager for the Always-On Relational Agent. We will first describe the structure of the agent and its components, then detail the workings of the Relationship Manager's planning and relevance algorithms. These algorithms implement the models from Chapter 3.

4.1 System Architecture

The agent has three main parts, as shown in Figure : the Relationship Manager, which makes plans on the scale of hours and days; the Collaboration Manager, which handles real-time interactions and various aspects of establishing and maintaining engagement, and also contains a Disco collaborative dialogure engine (Rich & Sidner 2012); and the shared components of the agent, including activity plugins and a semantic network of generally-used knowledge about the world and the user.

Disco is a program used to reason about hierarchical task networks and the focus of attention in dialogues. It executes instructions for performing specific activities, and also has meta-information about how tasks relate to each other, including preconditions and postconditions, orderings, or choices between options. Disco provides powerful methods for managing collaborative dialogue, planning, and inferring intent from action. It supports many of the capabilities of the Collaboration Manager.

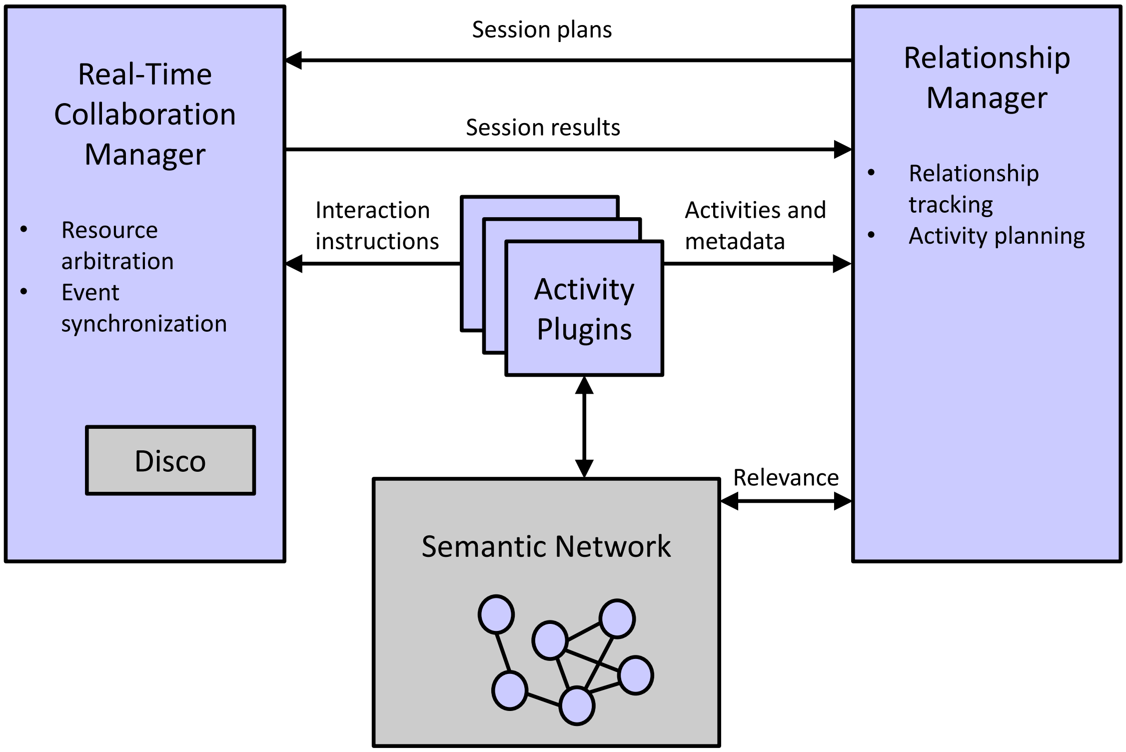

Figure 6 – System Architecture and Communication

The Collaboration Manager is the "face" of the system. It controls real-time concerns: establishing and maintaining engagement; executing plans provided by the Relationship Manager and plugins; selecting and interspersing "contextual" tasks (e.g., card games) that can be performed in parallel with other activities, such as conversations; and other duties to do with the proper functioning of the agent on short time scales. It utilizes Disco to support reasoning about task decompositions, intent behind certain actions, and the like.

The activity plugins contain domain-specific knowledge in the form of task models (providing instructions on how to perform the tasks in question), metadata, internal logic and interfaces, and the like. For instance, a "weather" plugin would have task models for discussing the weather, providing weather reports, and noting unusual weather patterns, along with metadata on these topics, all connected to code for retrieving raw forecast data from internet sources. Plugins read stored information about the world, the user, and the relationship from the other components of the agent (chiefly the Relationship Manager and the semantic network), and can adapt their behavior based on these factors, potentially altering which activities are available or how activities are performed. For example, the agent might address the user with greater formality and politeness when they are at a less familiar relationship stage. The activities encoded by plugins' task models are often purely conversational, i.e. discussions about various topics, but can also include activities such as playing games or helping arrange a schedule. In these cases, task models provide the code that allows the agent to perform specialized functions as required by the activity.

The semantic network is based on a standard ontology implemented using the Web Ontology Language (OWL) (McGuiness and Van Harmelen 2004), which includes some basic knowledge expected to be generally useful to the two Managers and the plugins. Abstractly, it contains information for reasoning about about familial relationships and locations, and concretely, it stores information about specific relations and friends of the human, their locations, and so forth. For instance, the network might include the user's sister and, adjacent to this node, the city where she lives. Any part of the agent can access this network and reason using its underlying ontology – determining that certain people live near each other, or are related, for example. The weather plugin mentioned above might use this network to find locations of particular interest to the human, such as the place where they live and the places where their immediate family live, and then provide information about unusual weather patterns in these areas. As will be described in Section 4.3.3, this semantic network is used as a topic relevance graph to implement the relevance algorithm described in Section 3.5.

The Relationship Manager handles longer-term aspects of interaction with the human. It tracks the progress of the relationship and provides information on the current relationship stage to other components. Its central duty is planning activities for interaction sessions.

4.2 Relationship Stages

As discussed in Chapter 3, we are using a model of relationship stages that delineate the set of activities available at each stage. For purposes of the Always-On Agent, we have a fairly restricted relationship stage lattice. There are three stages (*strangers*, *acquaintances*, and *companions*) arranged in a total order (see Figure 7). Thus, the relationship progresses from strangers to acquaintances to companions over some period of time. The model is therefore as follows. Strangers are permitted only simple, non-intimate interactions, with only a few activities available, and have low levels of closeness. Strangers may have some activities available that are not typical of other stages, e.g. introductions, but most permitted activities are not lost as the relationship progresses. Acquaintances are closer than strangers, and have a wider range of acceptable activities. A virtual agent that lives in your home is likely to progress to this stage fairly quickly. Companions are closer than either strangers or acquaintances, and are permitted the widest range of potential activities. It will probably take several weeks before this stage is reached for a given relationship.

Progression between these relationship stages occurs by a process of accumulating “social capital”. As described in Section 4.3.1, activities are assigned a quantity of “social utility”, the value of the activity as a social interaction. This value, in addition to its other uses, is added to a running total, representing the extent of social interactions between the user and the agent throughout the course of the relationship. The relationship progresses to a new stage when a certain threshold level of accumulated social utility is reached.

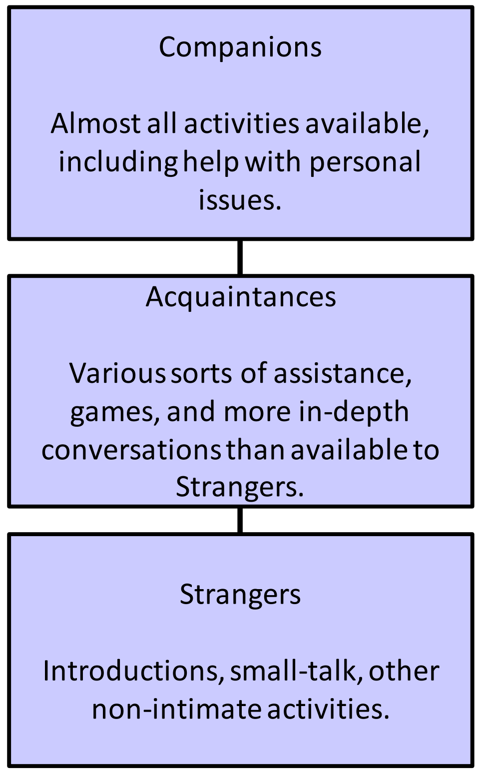

Figure 7 – Always-On Agent Relationship Stage Lattice

4.3 Planning Algorithm

The planning algorithm encompasses several steps, and includes implementations of the mathematical models described in Chapter 3.

4.3.1 Activities, Metadata, and Utility

To begin planning, the Relationship Manager requests a list of available activities from each of the activity plugins (see **Error! Reference source not found.**). Plugins determine which activities are available according to the relationship stage and their own internal logic – for example, "talk about the weather" would be available in any relationship stage, while "introduce myself" would be only available to strangers (except under quite unusual circumstances). The Relationship Manager then has a list of activities along with some important metadata about each one – for each individual activity (and a plugin may provide several activities, simultaneously or under different circumstances), the plugin governing that activity provides its social utility, its virtuous utility, its expected duration, and its required closeness. See Appendix B for example metadata of activities used in example planning scenario (Chapter 5) as well as the evaluation study (Chapter 6).

The social utility is a measure of the value of the task as a social activity, its benefit as an interaction, as well as the degree to which it builds up a level of closeness and intimacy during a particular interaction. The virtuous utility is the measure of the value of the task for other reasons – this depends on the Agent's design and the plugin in question, but it could encompass exercise, successfully completed tax returns, or improved cognitive function. In the Always-On Relational Agent, virtuous utility is derived from activities that help the human with organizational tasks, or that encourage exercise or external socializing and other activities. The total utility of a task is given by the sum of its social and virtuous utilities, as described in Chapter 3:

$$\boldsymbol{U}(\boldsymbol{A}) = \boldsymbol{S_D}\boldsymbol{S}(\boldsymbol{A}) + \boldsymbol{V_D}\boldsymbol{V}(\boldsymbol{A}) \qquad \text{(Eq. 1)}$$

The required closeness of an activity is a measure of how difficult it is to access, how carefully it must be broached. It is the amount of temporary closeness required for that task to be not merely *available*, but *accessible* (see Section 3.2) at a given moment. Temporary closeness starts at a baseline determined by the overall closeness of a relationship, and increments by the social utility of activities performed together. It fades over time, such that it is generally at the baseline rate at the beginning of a session, unless two sessions occur in close proximity.

The utility of an action at a given time is a dynamic value. Naturally, the utility of an action can be changed as desired by the internal logic of its governing plugin. Moreover, an action is worth less marginal utility if performed multiple times in rapid succession, so a damping effect is applied to the utility of individual actions (see Section 3.4). This fades over time in the same manner as does temporary closeness.

Finally, a larger-scale but weaker damping is applied to each kind of utility separately – *all* utilities of a given kind, virtuous or social, are reduced by a factor, the effect becoming stronger the greater the utility of that kind assigned to planned activities (see Section 3.4). In this manner, the agent avoids focusing too strongly on activities of a single kind.

4.3.2 Plan Construction

To begin planning, the Relationship Manager constructs a tree of all possible plans that fit within a given time window (see Figure 8a). On this first pass, the plan includes all activities that are possible to access given their position in the plan. The Collaboration Manager may select additional contextual activities to perform in parallel with those given in the plan (for example, playing a card game while discussing a recent baseball game). These activities may provide social utility not accounted for by activities occurring in the plan. Therefore, the Relationship Manager includes all activities that would be accessible given the amount of temporary closeness generated by the prior activities in the plan, plus an amount equal to the maximum that could possibly be generated by contextual activities by that point. The plan is limited to a specific duration, so when no activities that fit in the remaining time are accessible from a given point in the plan, no further activities are added at that point.

Next, the total utility of each path in the tree is calculated, with the utility of each activity evaluated in the context of the preceding activities as described above. From this, we derive the maximum possible utility, and take 90% of that value as the minimum allowable utility, so as to avoid unusual sensitivity and allow a variety of options in the plan. Figure 8a shows an example of a plan tree with path utilities calculated. Finally, the tree is pruned of all branches that contribute only to paths with total utilities below the allowable minimum. Thus, the resulting tree is such that any path from the root to a leaf produces a plan with a utility close to the maximum possible. The Collaboration Manager may therefore allow users choice in determining the course of a given interaction.

Figure b is an example of a resulting plan tree, involving three activities: discussing baseball, discussing the weather, and organizing a schedule. An interaction would proceed from left to right, starting at either Baseball or Weather, and continuing to a leaf. Thus, there are four possible interactions corresponding to this tree – one for each leaf, or, equivalently, one for each path through the tree: Baseball, Scheduling; Baseball, Weather, Scheduling; Weather, Scheduling; and Weather, Baseball, Scheduling.

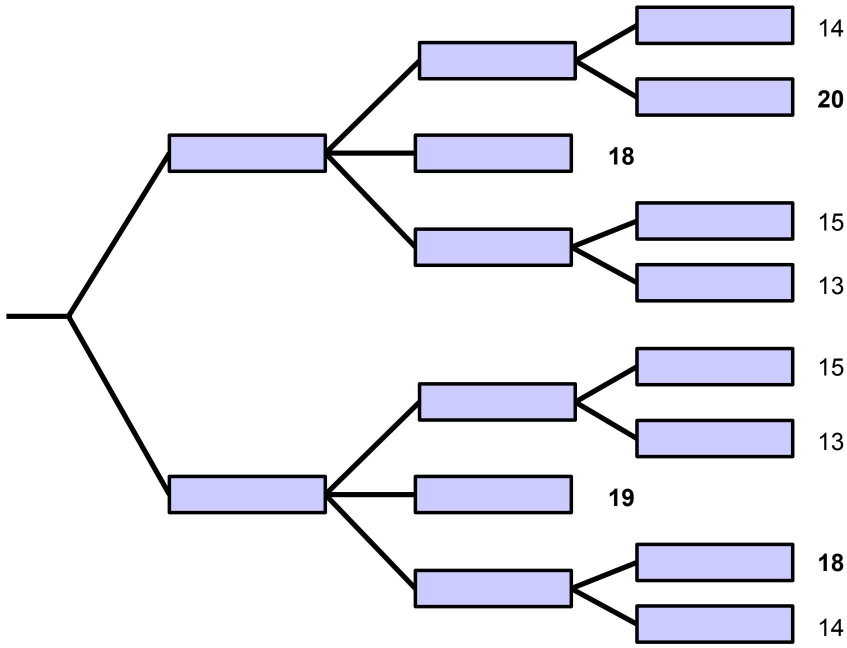

Figure 8a – Full Plan Tree

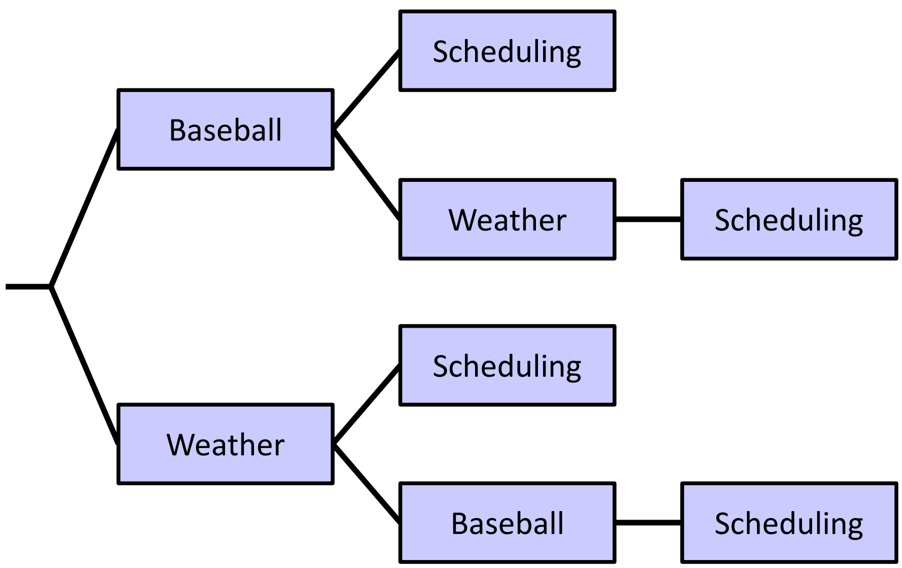

Figure 8b – Pruned Plan Tree

4.3.3 Relevance Algorithm

Before sending the completed plan tree to the Collaboration Manager, the different options at each junction are sorted according to topic relevance. As all paths are of approximately equal utility, and the system is an approximation of what is ultimately a 'fuzzy,' qualitative judgment, we ignore calculated utility in determining priority at this stage. Instead, we prioritize based on topical relevance. This process works as described in Section 3.5, using our semantic network, described above, as a topical relevance graph. Activities may be associated with one or more nodes of the graph, as determined by their governing plugins when they are sent to the Relationship Manager. For instance, a discussion about a person is associated with the node representing that person, while a discussion about the weather is associated with the location of the weather patterns under discussion. At the point in a plan after a given activity is completed, we consider all nodes associated with the activity, and all nodes adjacent to those, to be "active" – subsequent activities associated with active nodes are prioritized over those not so associated.

This process does not remove branches from the plan. Instead, it prioritizes amongst activities whenever there are multiple options. The Collaboration Manager then may emphasize higher-priority activities when presenting choices to the user.

As an example, a weather plugin might mention a snowstorm occurring in Florida if the user has a relative living there (as this would be rather unusual weather for Florida). In this case, the weather discussion would be associated with the "Florida" node of the semantic network (see Figure), and this node would be adjacent to the relative in question (as their place of residence). Thus, both of these nodes would be "active", and activities associated with either, such as displaying a photo album that this relative has recently posted to a social network, would have higher priority than unrelated activities.

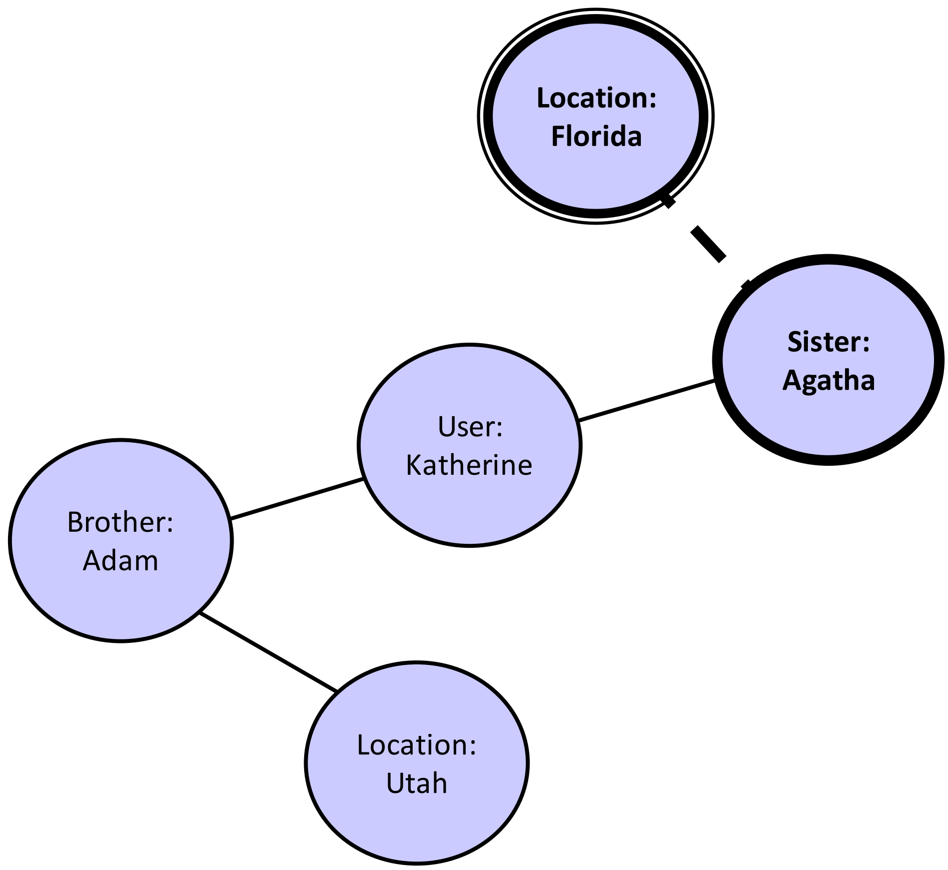

Figure 9 – "Snow in Florida" Example

4.3.4 Plan Execution

The plan is then sent to the Collaboration Manager. It executes the plan using task models provided by the activity plugins. The details of execution, and any parallel activities that may be added, are decided by the Collaboration Manager without input from the Relationship Manager. After each activity, the Collaboration Manager is able to provide multiple options, so the actual interaction can be steered largely by the user. As a result of the pruning process, any sequence of choices results in an acceptable interaction.

5. Example Scenario

To demonstrate the algorithms and models discussed in the previous two chapters, we will provide a detailed example scenario, in which we will work through the processes that occur in a sample run of the system. This description is taken from a running implementation of the Relationship Manager.

5.1 Overview and Context

In this scenario, the agent and human have been in the acquaintances stage of their relationship for some time, and are on the cusp of transitioning to the companions stage. The agent plans the first session of the day, proposing several activities and expecting to become companions by the end of the session. The Relationship Manager sends this plan to the Collaboration Manager, which executes the interaction itself. As it turns out, this attempt to advance the relationship is successful, but the user rejects all further proposed activities partway through the plan. The Collaboration Manager requests the Relationship Manager to make a new plan based on this new situation, and the Relationship Manager uses the established conversational context and the new relationship stage to come up with another session plan.

For purposes of this run-through, we will be using the set of example activities shown in Appendix B, selected to provide a range of alternatives for the planner, spanning different relationship stages and levels of closeness.

First, we set up the context for the planner: the current relationship stage is acquaintances, with a base temporary closeness of 3. There has been no preceding session on this day, so the current temporary closeness is at the base level. The Collaboration Manager requests a plan from the Relationship Manager.

5.2 Initial Plan Construction

The Relationship Manager requests available activities from the activity plugins. The plugins are able to access information about the relationship stage, and they use this along with their own internal logic to determine which activities to provide. In this case, the plugins provide all activities available to acquaintances. If the relationship had been one between strangers, a smaller set of activities would be available (ones that require less intimacy) but one additional activity would be provided in that case only: i.e., self-introduction, in which the agent introduces itself and provides some basic information about its functionality. Self-introduction is only available to strangers, and only once (though of course there may be other conversations about the agent's functionality available at other times).

Next, the Relationship Manager constructs the exhaustive plan tree. All activities requiring no more than a temporary closeness of 3 are selected as children of the root node, for each of these all activities requiring only the temporary closeness that would be accrued after that first activity are selected as children, and so forth for the entirety of the allowed duration. After this, the total utility of each path through the tree is calculated, and all branches of utility below 90% of the maximum are pruned off, leaving only high-utility plans. Finally, the topics relevant to each activity are compared, and activities with relevant topics adjacent in the agent's world ontology to those relevant to the prior activity are prioritized and placed higher on the list of possible subsequent activities. The final plan tree is as pictured in Figure .

Note that this process can produce plan trees in which different paths through the tree result in very different interactions. For instance, here, if the conversation turns to personal anecdotes after the baseball discussion, scheduling and medication will never be brought up in this session, due to the limited duration available. As all paths through the tree are calculated to have high utility, these options are all acceptable, and of course these alternate activities will likely still be available during the next planning session. For some activities, however, this might seem problematic. For instance, discussing medication might be optional at the moment, but crucial at some point during the week. This kind of judgment is performed by the activity plugins, rather than the Relationship Manager: as noted in Section 3.4, different activities will change utility in different ways if repeatedly postponed. In this case, if the medication discussion is not selected, the governing plugin of this activity might update the metadata it sends to the planner tomorrow, assigning a higher utility to the discussion.

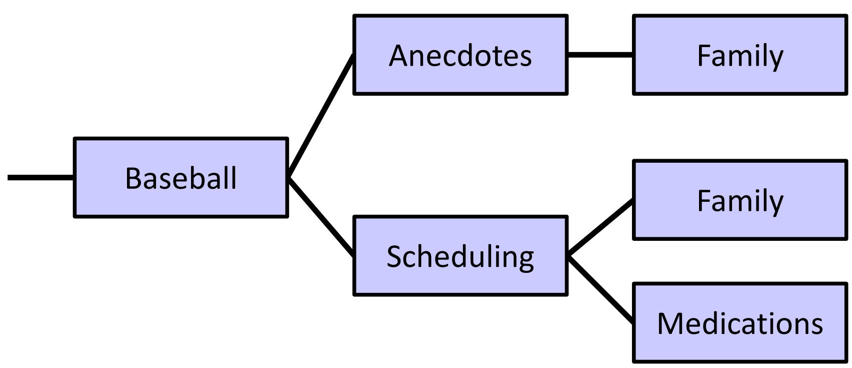

Figure 10 – Example Scenario: Initial Plan

5.3 Plan Execution

This plan tree is then delivered to the Collaboration Manager. The specifics of the Collaboration Manager's inner workings are beyond the scope of this thesis, so we shall simply describe the generalities. The Collaboration Manager at this point establishes engagement with the human, and begins the conversation. They discuss a recent baseball game, then work on arranging the human's schedule. However, at this point the human is not interested in any of the subsequently offered tasks, and the conversation ends.

5.4 Re-Planning

The Collaboration Manager then hands off to the other portions of the system – to the governing plugins of activities that were under consideration, and to the Relationship Manager, which is at this point unaware of the status of the interaction. Plugins receive information about their activities that can be used to alter available activities and metadata provided to the Relationship Manager. The Relationship Manager receives the entire interaction history, which it uses to determine which activities were performed and which ignored, and calculates the current level of closeness. It is ready to begin planning again, with the closeness decreased at the time of planning by any substantial delay between sessions.

In this particular example, there is an additional consideration. The total amount of social utility required for the next stage of the relationship has been reached, so the Relationship Manager now decides that the system and user are now companions. Thus, when the next round of planning begins, there are several differences in how things proceed. First, when the plugins receive the request to populate the list of available activities, there are several new activities provided in accordance with the advanced relationship stage. Next, the plan is contextualized by the immediately preceding session: there is some level of built-up temporary closeness, and, if within a brief window of time, the topic most recently at hand is still considered to determine the relative relevance of new activities with which to begin the next interaction.

The planning process is largely similar for this second iteration. Again, the planner constructs an exhaustive tree of all possible plans (given the new set of available activities), calculates the maximum utility, prunes away all non-optimal paths, and prioritizes based on topic relevance. However, in this round, activities performed and activities offered but rejected during the previous interaction are treated as having lower utility than in the initial plan. Finally, the new plan is sent back to the Collaboration Manager, to continue the interaction. This new plan is shown in Figure 4.

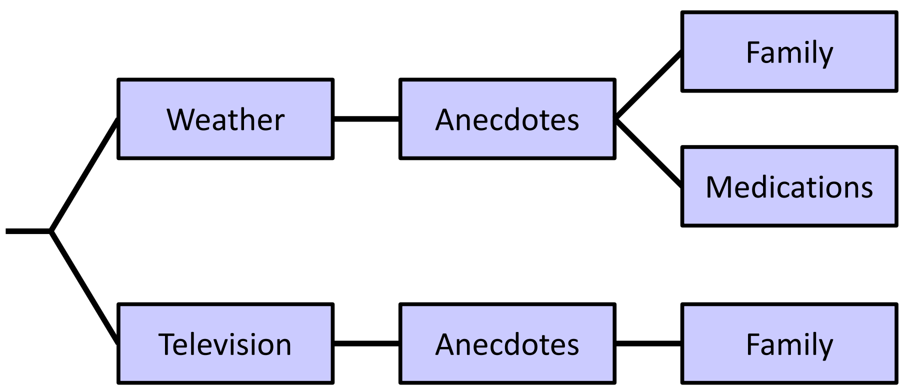

Figure 4 – Example Scenario: Second Plan

6. Evaluation

In order to verify the plausibility of these models and their suitability for use in a companion agent, we performed the study described in this chapter. Our goal was to determine if the features of the model produce behavior that is consistent with the aims of the project, i.e., natural-seeming interactions and relationship progression. At this point, it is vital to assess the viability of the underlying model before implementing it in the complete agent, to ensure that the system will work as desired when complete.

The materials used in this study are attached in Appendices A and B.

6.1 Experimental Design

The domain of social interaction is subtle and can be difficult to assess. Moreover, the overall system is not yet complete, and so cannot be used to engage in actual conversations with humans. Therefore, we have focused the study on three key features of the planner: the selection of activities appropriate to the current relationship stage, the selection of activities based on social utility, and the selection of activities based on the required level of temporary closeness. To study these factors, we compared the planner described in Chapter 4 (the "social planner") to an "antisocial" planner, designed to explicitly violate these key principles. This allows us to compare the social planner, and therefore the model it is built on, against an alternative that clearly contrasts with the model's assumptions. The antisocial planner differs from the social planner in three ways. First, while the social planner selected only from activities that are appropriate to the current relationship stage, the antisocial planner selected from *all* activities in the domain. Second, while the social planner tried to engage in social activities, the antisocial planner *subtracted* the social utility of an activity from its total utility, therefore focusing on *less* socially valuable activities. Third, while the social planner respected temporary closeness requirements to access activities, the antisocial planner ignored these requirements and was *more* likely to select activities the more inappropriate they were in terms of required temporary closeness – specifically, the antisocial planner increased the utility of an activity by the (nonnegative) difference between the closeness required and the closeness available, so long as the required closeness was above the available closeness. Thus, the social planner had the following as a utility function:

$$\boldsymbol{U(A) = S_D S(A) + V_D V(A)} \qquad \text{(Eq. 1)}$$

Whereas the antisocial planner had as a utility function:

$$\boldsymbol{U(A) = V_D \mathrm{V}(\mathrm{A}) - S_D S(A) + max(0, R(A) - C)} \qquad \text{(Eq. 7)}$$

Where *A* is an activity, *U(A)* is the total utility of an activity, *V(A)* is the *a priori* virtuous utility, *S(A)* is the *a priori* social utility, *R(A)* is the required closeness, and *C* is the current closeness.

The purpose of using the antisocial planner, and opposing the social planner in these three key areas, is to show the effects of the model as clearly as possible – by violating its key tenets this strongly, we are able to see whether these assumptions are effective in producing desirable behavior. Specifically, we are able to see whether, in aggregate, it is valuable to selectively screen off activities based on the perceived relationship stage, to emphasize social interaction, and to approach delicate topics more carefully at earlier stages. As we are chiefly attempting to determine whether the model as a whole is effective, it is difficult to disentangle these effects. One possible line of future inquiry would be to selectively ablate different components of the social planner, in order to determine to which degree each assumption contributes to the system's effectiveness.

As we tested specific scenarios rather than branching plans which could result in any of a number of different scenarios, the relevance algorithm was not tested in this evaluation.

Between the social planner, based on the model described in the preceding chapters, and the antisocial planner, which directly opposes this model, we hypothesized that the social planner will produce more natural social interactions, that is, ones that are more appropriate to a given relationship stage. We determined "naturalness" by the intuitive judgment of untrained judges (essentially the same metric by which normal interactions between humans are judged in day-to-day life). We expect judges to find scenarios planned by the social planner to be more natural-seeming and plausible than those produced by the antisocial planner. More precisely, based on the three key principles, our hypothesis is that the difference in naturalness between the results of the two planners will be greater the more these principles the social and antisocial planners to differ. This will occur when the antisocial planner is able to select activities unavailable at a given stage to the social planner, when a variety of high-social-utility activities are available to both planners, and when activities requiring high temporary closeness are available. To more clearly see the effects of the model, we compare the social planner to an antisocial planner that differs in all of the relevant ways simultaneously. This amplifies the differences between the planners, making the influence of the model more visible, but makes it more difficult to disentangle the influence of separate components of the model. However, the above factors are most strongly in play overall at the earlier relationship stages, when the social planner is selecting activities from a much more restricted set than the antisocial planner, so we expect to see a stronger effect the less familiar the relationship.

To compare these planners, we required contexts for them to plan within. Most broadly, we established that the interactions would be presented as occurring between an older adult living alone (the target audience of the current Always-On Relational Agent) and a community worker who has volunteered to provide some companionship and assistance. We selected this context so that study participants would be able to judge the naturalness of the interaction in an intuitive manner, rather than forcing them to attempt to extrapolate what a "natural" interaction should be between a human and an artificially intelligent agent. The community worker role was selected specifically because it is broadly similar in function and duties to the Always-On Relational Agent, and in fact the agent is in many ways patterned after community workers working with the elderly. We used the same set of example activities as introduced in Chapter 5 (see Appendix B), and specified three different relationship conditions – one for each of the three relationship stages, strangers, acquaintances, and companions. Thus, we have six different conditions for the models, one of each relationship stage for each of the two different planner utility functions. To compare them, we had the planner produce two high-utility plans for each condition (selected to be as distinct as possible while retaining high utility), and grouped them by relationship stage. Therefore, for each stage, we have four different plan scenarios – two each from the social and antisocial utility functions. Study participants can then compare the relative naturalness of the plans for each scenario.

As a method of rating these scenarios, participants were asked to construct partial orders, ranking the different options relative to each other in terms of their degree of naturalness and plausibility – that is, the degree to which these interactions seemed as though they would reasonably occur in an interaction between two people with the relationship under consideration.

6.2 Experimental Procedure

For this study, we had 12 participants (5 female, 7 male), each an enrolled student at Worcester Polytechnic Institute. Participants were compensated with either $5 or credit for participation in research (required for sociology students). Each participant was read a description of the scoring system, then presented with a sample sheet (the study materials are attached in Appendix A), formatted similarly to the sheets to be used in the study proper, describing a set of meals and requesting the participant to rate them according to their degree of healthfulness. The partial order rankings for these sheets were not examined – the purpose of the sheet was solely to familiarize the participants with the scoring system.

After this preparation, participants were provided with the context for the experiment. They received a description of the two people putatively involved in the interactions: Katherine, a socially isolated older adult, and Samantha, a community worker who has volunteered to visit Katherine and provide some aid. Additionally, they received an explanation of the criteria upon which the scenarios were to be judged. Next, participants received their first sheet of scenarios to compare. Each following sheet was provided upon completion of the previous one. Each participant received one sheet for each of the three relationship stages, in a random order. Each sheet contained a description of the relationship stage, along with a description of each of the four corresponding plans, also in a random order. For each sheet, they were requested to rank the plans in a partial order according to their relative naturalness and plausibility.

6.3 Results and Discussion

From this procedure, we received a number of partial order rankings of the various scenarios for each of the relationship stages. For each of the partial orders, we extracted a measure of the preference between the two planners by determining the number of times a social planner scenario was ranked above an antisocial planner scenario and the number of times the converse occurred. We used standard sign tests on this data to determine the strength of the preference, comparing the relative numbers of positive events (a social planner scenario being ranked above an antisocial planner scenario) and negative events (an antisocial planner scenario being ranked above a social planner scenario).

Strangers	Acquaintances	Companions	Social > Anti	Social < Anti
1 2 / 3 4	1 2 / 3 / 4	4 / 3 / 1 / 2	8	4
1 / 3 / 2 / 4	1 / 2 4 / 3	2 / 4 / 1 3	8	2
1 / 2 / 3 / 4	2 / 1 / 3 4	1 2 / 3 / 4	12	0
1 / 2 / 3 / 4	2 / 1 / 4 / 3	4 / 1 / 2 / 3	10	2
1 / 3 / 2 / 4	3 / 2 / 4 / 1	2 / 1 / 3 4	8	4
1 / 2 / 3 / 4	2 / 1 / 3 / 4	4 / 2 / 1 / 3	10	2
1 2 / 3 / 4	3 / 1 2 / 4	3 4 / 1 2	6	6
1 2 / 3 4	1 / 2 / 3 / 4	2 4 / 1 3	9	1
1 / 3 / 2 4	1 / 2 4 / 3	1 2 / 4 / 3	9	1
3 / 1 2 / 4	2 / 1 4 / 3	2 / 1 4 / 3	8	2
1 / 2 3 4	2 / 1 4 / 3	1 2 / 3 4	9	0
2 / 1 / 3 / 4	2 / 1 3 / 4	4 / 2 / 3 / 1	8	3
Total Positive / Negative per Stage			Overall Totals	
40 / 5	38 / 5	27 / 17	105	27

Figure 5 – Evaluation Results

Figure 5 shows the collected results of the study. The first three columns show the precise partial order rankings given, where 1 and 2 represent scenarios generated by the social planner, and 3 and 4 represent scenarios generated by the antisocial planner. Partial orders are given from most highly ranked scenarios to least, with equivalence classes separated by slashes. Columns 4 and 5 give the number of occurrences in a given row of a social planner scenario being ranked higher than an antisocial planner scenario (positive results), and of an antisocial planner scenario being ranked higher than a social planner scenario (negative results), respectively. The final row of the table gives total counts of these occurrences for each relationship stage, as well as overall totals.

In Appendix A we have included the sheets of scenarios for each relationship stage, with the scenarios of each sheet ordered (1 2 3 4) – two social-planner scenarios, labeled A and B, followed by two antisocial-planner scenarios, labeled C and D. For an example of how this encoding system works, the first entry in the “Strangers” column of Figure 5 represents a sheet in which the participant wrote A and B in one box, and C and D in a lower box.

Overall, there were 105 positive results and 27 negative results, giving a p-value of 2.07e-11. The social planner therefore overall produced plans that were more natural than those of the antisocial planner, with high significance. The result was also strongly significant for the specific cases of the strangers and acquaintances relationship stages, with respective p-values of 7.88e-8 and 2.5e-7. However, in the case of the companions relationship stage, the result was not significant (p-value 0.174).

That a much weaker correlation was found for the companions relationship than for the less close relationships is plausible on both psychological and algorithmic grounds. On the one hand, behavior that deviates from proper social standards is more easily tolerated from those we are closer with. On the other hand, one of the three rules violated by the antisocial planner was the restricted availability of certain activities depending on the relationship stage, and for the companions relationship all activities were available to the social planner was well – thus, both planners were selecting from the same pool of activities for this stage alone.

These results suggest that the algorithms we developed are capable of serving as the basis for a social planner that can produce plausible recommendations for natural interactions. Moreover, they suggest that, at least to an outside observer, the most salient function of the planner is the higher-level judgment of whether an activity should be considered available at all, given the current relationship.

7. Conclusion and Future Work

As social AI becomes more advanced, the expectations that people place on robots and virtual agents will expand, encompassing many of the features of human interaction that we take for granted. These agents will be assumed and expected to have more and more of the basic capabilities that humans use for navigating social situations. These are extremely complex capabilities that are largely unconscious, and therefore both deceptively simple-seeming and difficult to explicitly address. Relationship modeling is essentially about providing a basis for systematically addressing these expectations, and this research serves as a starting point for the development of robust models of the intricacies of relational development and interaction.

The contribution of this work in the larger context of relationship modeling is twofold. First, it provides a concrete model and implementation for an advanced relational agent. In addition to the explicit work done by the model in selecting activities and planning natural-seeming interactions, we have also provided a framework within which content can be developed so as to take advantage of information about relationships. This relationship information could inform the formality of discussions, the aggressiveness of game AIs, or the degree of qualification and explanation needed during activities. Second, this work provides an implemented and evaluated foundation for the development of more advanced models and agents. These models could utilize the contextual information noted above, as well as the mathematical models of relationship progression, or attempt to solve similar or different problems using novel strategies, with the current model as a point of comparison.

While the current models are useful for the Always-On Relational Agent, there are some clear directions in which further research could take this work. The Always-On Relational Agent is still under development, and currently involves research that may motivate alterations to the agent and its Relationship Manager. Vardoulakis et al. (2012), as part of the extended Always-On Relational Agent project, have conducted a Wizard-of-Oz study of interactions between humans and virtual agents. This study will inform the further development of the project, and the ways in which the Relationship Manager and its underlying models will change and be used. Further afield, more work remains to be done experimentally testing the many parameters of the model, and developing competing models to determine which provides the most predictive, explanatory, and natural-behavior-generating power.

More specifically, there are several natural areas of further inquiry that arise from the current research. It would be useful to perform a study with selective ablation of the model's components, to determine the effects of each part. Such studies could also compare the model to alternative planning systems, or examples of real interactions planned by humans. Further studies could extend the model or otherwise study novel aspects of relationships. For instance, a study could investigate the ending of a relationship, and the ways the relationship model could guide this process.

Social AI, the art and science of precisely describing and dissecting the innumerable complex facets of human social interaction, is yet in its infancy. The development of this field requires assigning numbers and equations to fuzzy, qualitative processes, capturing the essence of intricate phenomena without excluding any of the myriad features of interactions that seem insignificant or even unnoticeable in the abstract, but whose absence will be sorely noted. This task will require drawing on the work of many disciplines, from psychology to sociology to neuroscience. Ultimately, it will involve simultaneously discovering and engineering the positions that artificial intelligences will take in our society. This work is designed to help us along that path.

Appendix A: Study Materials

This appendix includes the materials used in the evaluation study: the instructions read to the participants, the sample "healthy meals" sheet for demonstrating and practicing the partial order ranking system, three sheets of social planning scenarios, and the debriefing sheet. The planning scenario sheets represent only a single permutation of the scenarios – the first two scenarios on each sheet were generated by the social planner, the other two by the antisocial planner. In the study, four permutations of each sheet were used, identical to these save for the order in which the scenarios were printed. The sheets here represent the permutation (1 2 3 4), the other permutations used being (2 4 1 3), (3 1 4 2), and (4 3 2 1).

Instructions (read to participant/do not give)

In this study, you will be presented with written descriptions of different situations and you will be asked to rank them according to a specified criterion.

First we will practice how to fill out the ranking forms, using an example that is not part of this study.

[Hand practice sheet to participant.]

Please read each option, marked A, B, C, and D, then write each letter once in the ranking boxes on the left of the form. Write the letter or letters for the best option or options in the top box, the second best in the second box, and so forth. You may put more than one letter in the same box if you feel they are equally good options (in which case some of the boxes may be empty when you are done).

[Let participant fill in form.]

Do you understand how to use this ranking form?

Do you have any questions about it?

[Next page]

Each of the following options is a meal consisting of several different foods.

Please rank the four meals below in terms of how healthy you think they are by writing each label (A, B, C, and D) in an appropriate box on the left. You may put more than one label in a box (and leave some boxes empty) if you think some choices are tied.

Most Healthy

Less Healthy

A

A bottle of soda.
A cheeseburger with bacon.
French fries.
Two scoops of ice cream.

B

A glass of milk.
A roast chicken sandwich.
A mixed green salad.
An apple.

C

A glass of water.
A slice of white bread.
A small piece of cheese.
Some green beans.

D

A mug of coffee.
A slice of pizza.
A cup of instant noodles.
A bar of chocolate.

Now we move on to the main study.

In this study you will be reading and ranking descriptions of social interactions between two people in terms of how natural and plausible you think they are. The two people are:

* **Katherine** is an older adult living alone

* **Samantha** is a community worker who has volunteered to visit Katherine

You will have three forms to fill out, which are organized just like the practice form. Each form has four different interaction descriptions labeled A, B, C, and D for you to rank.

For each form, you will be asked to imagine a different social relationship between the two people. Consider each form separately and independently. Please rank the options based on how **natural** the interaction seems, i.e., how **plausible** it would be for the two people to have the described interaction, given what you know about their social relationship.

Do you have any questions before I give you the first form to fill out?

[Give first form]

Let me know when you are done. Please take your time, as you only have three forms to fill out for this study.

[Give other forms in turn.]

Thank you for participating in the study. Do you have any comments or questions?

[Write down comments/questions]

In these scenarios, Samantha and Katherine have only met once before, and are not yet familiar with each other.

Please rank the four interactions on the diagram below in terms of how natural and plausible they seem, using the labels A, B, C, and D.

Most Natural/Plausible

Less Natural/Plausible

A

First, Samantha and Katherine talk about the basic services a community worker provides for 20 minutes.
Next, they talk about a recent baseball game for 20 minutes.
After that, they talk about the weather for 20 minutes.
Then, they watch TV together for 20 minutes.
Next, they trade humorous anecdotes for 20 minutes.
Finally, they discuss the mayoral election for 40 minutes.

B

First, Samantha and Katherine play a card game together for 40 minutes.
Next, they discuss the mayoral election for 40 minutes.
After that, they trade humorous anecdotes for 20 minutes.
Then, they talk about a recent baseball game for 20 minutes.
Finally, they talk about the basic services a community worker provides for 20 minutes.

C

First, Samantha and Katherine sort out Katherine's medicine for 20 minutes.
Next, Samantha takes 20 minutes to convince Katherine to write a letter she's been putting off.
After that, they talk about Katherine's recent diagnosis with a serious illness for 40 minutes.
Then, they discuss the mayoral election for 40 minutes.
Finally, they organize Katherine's schedule for 20 minutes.

D

First, Samantha takes 20 minutes to convince Katherine to write a letter she's been putting off.
Next, they discuss Katherine's brother's illness for 40 minutes.
After that, they talk about Katherine's recent diagnosis with a serious illness for 40 minutes.
Then, they sort out Katherine's medicine for 20 minutes.
Finally, they organize Katherine's schedule for 20 minutes.

In these scenarios, Samantha and Katherine have met several times over the past two weeks, and have become somewhat comfortable with each other.

Please rank the four interactions on the diagram below in terms of how natural and plausible they seem, using the labels A, B, C, and D.

Most Natural/Plausible

Less Natural/Plausible

A

First, Samantha and Katherine talk about a recent baseball game for 20 minutes.
Next, they trade humorous anecdotes for 20 minutes.
After that, Katherine tells some stories from her childhood for 40 minutes.
Then, they organize Katherine's schedule for 20 minutes.
Next, they talk about Katherine's family for 20 minutes.
Finally, Samantha helps organize Katherine's medicine for 20 minutes.

B

First, Samantha and Katherine talk about a recent baseball game for 20 minutes.
Next, they organize Katherine's schedule for 20 minutes.
After that, Samantha helps organize Katherine's medicine for 20 minutes.
Then, they talk about Katherine's family for 20 minutes.
Next, Katherine tells some stories from her childhood for 40 minutes.
Finally, they trade humorous anecdotes for 20 minutes.

C

First, Samantha helps organize Katherine's medicine for 20 minutes.
Next, Samantha takes 20 minutes to convince Katherine to write a letter she's been putting off.
After that, they discuss Katherine's brother's illness for 40 minutes.
Then, they organize Katherine's schedule for 20 minutes.
Finally, they talk about Katherine's recent diagnosis with a serious illness for 40 minutes.

D

First, Samantha takes 20 minutes to convince Katherine to write a letter she's been putting off.
Next, they talk about Katherine's recent diagnosis with a serious illness for 40 minutes.
After that, Samantha helps organize Katherine's medicine for 20 minutes.
Finally, they do an exercise routine together for 60 minutes.

In these scenarios, Samantha and Katherine <u>have been meeting frequently for several months</u>, and have become very close.

Please rank the four interactions on the diagram below in terms of how natural and plausible they seem, using the labels A, B, C, and D.

Most Natural/Plausible

Less Natural/Plausible

A

First, Samantha and Katherine talk about a recent baseball game for 20 minutes.
Next, they organize Katherine's schedule for 20 minutes.
After that, they talk about Katherine's family for 20 minutes.
Then, they sort out Katherine's medicine for 20 minutes.
Next, Samantha takes 20 minutes to convince Katherine to write a letter she's been putting off.
Finally, they talk about Katherine's recent diagnosis with a serious illness for 40 minutes.

B

First, Samantha and Katherine talk about a recent baseball game for 20 minutes.
Next, they organize Katherine's schedule for 20 minutes.
After that, they sort out Katherine's medicine for 20 minutes.
Then, they talk about Katherine's recent diagnosis with a serious illness for 40 minutes.
Next, Samantha takes 20 minutes to convince Katherine to write a letter she's been putting off.
Finally, they talk about Katherine's family for 20 minutes.

C

First, Samantha takes 20 minutes to convince Katherine to write a letter she's been putting off.
Next, they sort out Katherine's medicine for 20 minutes.
After that, they discuss Katherine's brother's illness for 40 minutes.
Then, they organize Katherine's schedule for 20 minutes.
Finally, they talk about Katherine's recent diagnosis with a serious illness for 40 minutes.

D

First, Samantha takes 20 minutes to convince Katherine to write a letter she's been putting off.
Next, they talk about Katherine's recent diagnosis with a serious illness for 40 minutes.
After that, they discuss Katherine's brother's illness for 40 minutes.
Then, they sort out Katherine's medicine for 20 minutes.
Finally, they organize Katherine's schedule for 20 minutes.

Debriefing Form

Natural Interactions over Differing Relationship Stages

When you spend time with someone, the types of interactions you have with them change as you become closer - the sorts of conversations that strangers have are very different from those that good friends have. We have been developing a computational model of some aspects of this progression for use in an artificial intelligence, and the purpose of this study is to determine whether our model is actually recommending realistic, plausible behavior.

The scenarios you looked at were produced by two different processes - a "social" planner, which selected activities that were socially valuable and waited until an appropriate time to suggest certain activities, and an "antisocial" planner that deliberately violated these rules. We predict that the social planner will produce the most natural behavior, and therefore an artificial intelligence using this planner will be easier and more pleasant to interact with.

If you have any questions or comments, please ask me. If you have other questions or comments, please contact William Coon at [contact information removed]. Because other students may be participating in this study in the future, we ask that you not discuss the details of this experiment with your friends or classmates.

Thank you!

Appendix B: Example Activities

These activities, with attendant metadata, were used in the evaluation study (to generate the scenarios that the participants were requested to judge) as well as in the example planning scenario described in Chapter 5.

Activity Description	Social Utility	Virtuous Utility	Duration	Required Closeness
Self-introduction, description of basic services*	1	3	1	0
Talk about a recent baseball game	2	1	1	0
Talk about the weather	1	1	1	0
Play a card game	2	1	2	0
Watch television	1	0	1	0
Trade humorous anecdotes	2	1	1	2
Discuss a local mayoral election	2	2	2	2
Talk about favorite books	4	1	2	3
Listen to personal childhood stories	4	2	2	3
Organize schedule	1	3	1	3
Do an exercise routine	2	5	3	3
Talk about family	3	2	1	4
Sort out medications	1	3	1	4
Work on a scrapbook	3	2	2	5
Read emails from family, and write responses	3	2	2	5
Discuss a non-recent death in the family	6	1	2	7
Convince human to write an important letter	1	5	1	7
Discuss a sibling’s current illness	4	4	2	8
Discuss a recent serious diagnosis	3	6	2	8

* This activity is only available to Strangers who have not yet been introduced.

Appendix C: Plan Output

This is the output of the Relationship Manager in the scenario given in Chapter 5, in the form of a pair of Disco task models, encoding the plans given in Figures 10 and 11. These task models are formatted according to the ANSI/CEA-2018 standard (Rich 2009).

```
<taskModel xmlns=”http://ce.org/cea-2018”
about="urn:relationships.wpi.edu:examples:Today">
      <task id="Today">
            <subtasks id="decomp0">
                  <step name="Baseball" task="Baseball"/>
                  <step name="choice0" task="choice0"/>
            </subtasks>
      </task>
      <task id="choice0">
            <subtasks id="decomp1">
                  <step name="Anecdotes" task="Anecdotes"/>
                  <step name="choice1" task="choice1"/>
            </subtasks>
            <subtasks id="decomp3">
                  <step name="Schedule" task="Schedule"/>
                  <step name="choice2" task="choice2"/>
            </subtasks>
      </task>
      <task id="choice1">
            <subtasks id="decomp2">
                  <step name="Family" task="Family"/>
            </subtasks>
      </task>
      <task id="choice2">
            <subtasks id="decomp4">
                  <step name="Family" task="Family"/>
            </subtasks>
            <subtasks id="decomp5">
                  <step name="Pills" task="Pills"/>
            </subtasks>
      </task>
</taskModel>
```

```
<taskModel xmlns="http://ce.org/cea-2018"
about="urn:relationships.wpi.edu:examples:Today">
      <task id="Today">
            <subtasks id="decomp0">
                  <step name="Weather" task="Weather"/>
                  <step name="choice0" task="choice0"/>
            </subtasks>
            <subtasks id="decomp4">
                  <step name="Television" task="Television"/>
                  <step name="choice2" task="choice2"/>
            </subtasks>
      </task>
```

```
    <task id="choice0">
        <subtasks id="decomp1">
            <step name="Anecdotes" task="Anecdotes"/>
            <step name="choice1" task="choice1"/>
        </subtasks>
    </task>
    <task id="choice1">
        <subtasks id="decomp2">
            <step name="Family" task="Family"/>
        </subtasks>
        <subtasks id="decomp3">
            <step name="Pills" task="Pills"/>
        </subtasks>
    </task>
    <task id="choice2">
        <subtasks id="decomp5">
            <step name="Anecdotes" task="Anecdotes"/>
            <step name="choice3" task="choice3"/>
        </subtasks>
    </task>
    <task id="choice3">
        <subtasks id="decomp6">
            <step name="Family" task="Family"/>
        </subtasks>
    </task>
</taskModel>
```

Bibliography

Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. New York: Holt, Rinehart and Winston.

Bickmore, T. W. (2003). *Relational agents: Effecting change through human-computer relationships*.

Bickmore, T. W., and R. W. Picard (2005). “Establishing and Maintaining Long-Term Human-Computer Relationships”. *ACM Transactions on Computer-Human Interaction: a Publication of the Association for Computing Machinery.* 12 (2): 293.

Burgoon, J. K., & Hale, J. L. (September 01, 1984). “The fundamental topoi of relational communication.” *Communication Monographs, 51,* 3, 193-214.

Cole, T., & Bradac, J. J. (January 01, 1996). “A lay theory of relational satisfaction with best friends.” *Journal of Social and Personal Relationships, 13,* 1, 57-84.

Gilbertson, J. A., Dindia, K., & Allen, M. (December 01, 1998). “Relational continuity constructional units and the maintenance of relationships.” *Journal of Social and Personal Relationships, 15,* 6, 774-790.

Gratch, J., Okhmatovskaia, A., Lamothe, F., Marsella, S. C., Morales, M., van, . W. R. J., & Morency, L.-P. (January 01, 2006). “Virtual Rapport.” *Lecture Notes in Computer Science,* 4133, 14-27.

Hardin, Curtis D.; Conley, Terri D., Moskowitz, Gordon B. (Ed), (2001). “A relational approach to cognition: Shared experience and relationship affirmation in social cognition.” *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition.*, (pp. 3-17). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, viii, 503 pp.

Klamer, Tineke, and Somaya Ben Allouch (2010). “Zoomorphic robots used by elderly people at home.” *Proceedings of the 27th International Conference on Human Factors in Computing Systems. ACM International Conference on Human Factors in Computing Systems (CHI-2010).* Atlanta, GA: ACM Press.

Lim, T. (1994). “Facework and Interpersonal Relationships”. In S. Ting-Toomey (Ed.), *The challenge of facework: Cross-cultural and interpersonal issues*, (pp. 209-229). Albany, NY: State University of New York Press.

McGuiness, D., Van Harmelen, F., et al. (2004). “OWL web ontology language overview”. Technical report, W3C Recommendation.

Nickerson, R. S. (October 01, 2001). “The Projective Way of Knowing: A Useful Heuristic That Sometimes Misleads.” *Current Directions in Psychological Science, 10,* 5, 168-172.

Nooraei, B., Liu, Z., Conley, C., Sidner, C., Rich, C., and Bickmore, T. (September 2011). “Toward an Always-On Relational Agent for Social Support of Isolated Older Adults.” *GALA Video Festival of the Int. Conference on Intelligent Virtual Agents*, Reykjavik, Iceland.

Ochsner, K. N., & Lieberman, M. D. (September 01, 2001). “The Emergence of Social Cognitive Neuroscience.” *American Psychologist, 56,* 9, 717-34.

Rich, C. (2009), “Building Task-Based User Interfaces with ANSI/CEA-2018”, *IEEE Computer*, Vol 42, No. 8, August 2009, pp. 20-27.

Rich, C., Holroyd, A., Ponsler, B., and Sidner, C. (March 2010). “Recognizing Engagement in Human-Robot Inteaction.” *Proceedings of the 2010 ACM/IEEE International Conference on Human Robot Interaction*. Pp. 375-382.

Rich, C., and Sidner, C. (2012). “Using Collaborative Discourse Theory to Partially Automate Dialogue Tree Authoring”. *Int. Conf. on Intelligent Virtual Agents*, Santa Cruz, CA, Sept 2012.

Shibata, T., Kawaguchi, Y., & Wada, K. (January 01, 2012). “Investigation on People Living with Seal Robot at Home.” *International Journal of Social Robotics, 4,* 1, 53-63.

Stafford, L., Dainton, M., & Haas, S. (January 01, 2000). “Measuring Routine and Strategic Relational Maintenance: Scale Revision, Sex versus Gender Roles, and the Prediction of Relational Characteristics.” *Communication Monographs Virginia-, 67,* 306-323.

Turunen, M., J. Hakulinen, O. Stahl, B. Gamback, P. Hansen, M. C. Rodriguez Gancedo, R. S. de la Camara, C. Smith, D. Charlton, and M. Cavazza (2011). “Multimodal and mobile conversational Health and Fitness Comanions”. *Computer Speech & Language*. 25 (2): 192-209.

Vardoulakis, L. P., Ring, L., Barry, B., Sidner, C., and Bickmore, T. (2012). “Designing Relational Agents as Long Term Social Companions for Older Adults”. *Int. Conf. on Intelligent Virtual Agents*, Santa Cruz, CA, Sept. 2012.

Wish, M., Deutsch, M., & Kaplan, S. J. (January 01, 1976). “Perceived dimensions of interpersonal relations.” *Journal of Personality and Social Psychology, 33,* 4, 409-420.