Designing for Friendship: Becoming Friends with Your ECA

Bas Stronks, Anton Nijholt, Paul van der Vet & Dirk Heylen
Parlevink Research Group
Faculty of Computer Science
University Of Twente, PO Box 217
Enschede, The Netherlands
anijholt@cs.utwente.nl

"The imaginary friends I had as a kid dropped me because their friends thought I didn't exist."
Aaron Machado

1. Introduction

This paper is about the design of Embodied Conversational Agents (ECAs). In this field of Human Computer Interaction (HCI) and Artificial Intelligence (AI), the design of ECAs, or 'virtual humans', and the communication between those agents and human users is the main object of our research. A lot of effort is put into research to make ECAs more lifelike and believable and to make communication with ECAs more effective, efficient, and more fun. In order to do so, the idea was to make the agent more actively concerned with the relationship with the user. As a 'lay psychologist', we all know that people that you like (or your friends) are able to help you better, teach you better, and generally are more fun to interact with, than people that you don't like. However, 'liking' is person dependent. Not everybody likes the same person, and one person is not liked by everyone.

These observations sparked our interest in the application, effects, and design of a 'virtual friend'. An agent that observes it's user, and adapts it's personality, appearance and behavior according to the (implicit) likes and dislikes of the user, in order to 'become friends' with the user and create an affective interpersonal relationship. This agent might have additional benefits over a 'normal' Embodied Conversational Agent in areas such as computer assisted teaching and entertainment.

The above being the basis of our research, we were looking for a sound foundation on which to build research on a 'virtual friend'. This led us to the field of psychology. There is extensive knowledge about human interpersonal relationships in the field of personality and social psychology. However, psychology suffers from the fact that every person has 'judgment of character' and insight into 'human nature', which makes everybody a self proclaimed 'lay psychologist'. Hence, research and application of sound psychology research in other areas (such as computer science) is often found unnecessary. But lay psychology and folk wisdom is often proven wrong by psychology research. For example, folk wisdom dictates that, in human interpersonal relationships, 'opposites attract'. If a researcher would try to increase Human-ECA attraction by designing an ECA that is exactly the opposite of the human in terms of personality characteristics, it would be likely that the ECA would not be effective, because it has been shown hat attraction is greater when personality characteristics are similar. Thus, ECA design decisions need to be based on sound psychology research and not on 'folk wisdom' or otherwise.

Furthermore, recent computer science history already shows that application of rigid psychology in computer science is very useful: the application of cognitive psychology in computer science in the last two decades has resulted in improvements in problem-solving skills and task related behavior of computer users, thus improving Human-Computer Interaction effectiveness. Because communication with an ECA is social in nature, improving Human-ECA interaction should be accomplished by applying findings from social psychology. Consequently, a main aspect of our research is the application of (social) psychology in ECA design.

In this paper we first try to formulate some explicit research questions, showing the methodology in our research on designing for friendship (section 2). Then, in section 3 we discuss how to handle these questions. What do we mean by friendship and how can we exploit the tendency we see in humans to assign human properties to animals and objects when we interact with them? Section 4 is devoted to the possibility of translating the main aspects of human-human friendship to human-ECA friendship.

Obviously, we have to know which aspects play a role in the human-human friendship relationship. In section 5 we discuss how we can incorporate our findings in the design process, using a scenario-based design. We distinguish between the initial design of an ECA and the possibility to change the ECA characteristics according to an adaptation strategy based on knowledge obtained by interacting with a particular user. Before the conclusions (section 9) we have a short discussion about using stereotypes in designing ECAs and some related ethical observations in section 8.

2. Research Questions

The object of the research is to extend the knowledge in the field of Embodied Conversational Agent (ECA) technology in Human-Computer Interaction, concerning human-ECA relationships and human-ECA friendship, in order to make Human-Computer Interaction more effective. The research questions guiding this research are:

1. What is known about friendship in psychology research? In answering this question, we explored the psychology behind friendship, including:

    (a) What are the psychological mechanisms behind the formation of friendship?

    (b) What are variables and issues affecting friendship?

    (c) What are the effects of friendship on interaction?

2. How can we apply findings from research question 1 to Human-Computer Interaction? In answering this question, we took findings from psychology and translated them to a human-ECA situation. Sub questions were:

    (a) Is it possible to apply social psychology to human-ECA relationships, and if so, how?

    (b) How can we translate the knowledge about friendship in social psychology to useful insights in human-ECA friendship?

3. How does the answer to research question 2 affect the creation of ECAs? In this research question we applied he translation of human psychology to the design phase, design issues, and architecture of ECAs. Research sub questions were:

    (a) Given the translation of psychology, how can the design phase of an ECA be adapted to accommodate 'friendlier' ECA creation?

    (b) What are important issues in the design of 'friendly' ECAs and how can they be addressed?

    (c) Given what is known about human-ECA relations, how can the architecture of an ECA be adapted in order to create 'friendlier' ECAs?

3. Approach and Answers

Because of the novelty of the object of research and limited time, this research has an exploratory and theoretical character, and focus has not yet been on a possible implementation. Moreover, because of space limitations we can not address all the (sub-)questions separately and give all the details of our findings. See however [16]. Research question 1 has been answered by means of literature research in the psychology literature, specifically research in social psychology of interpersonal relationships and attraction, and has resulted in a description of relevant theory. We will make references to the results of this research in the forthcoming sections. The answer on research question 2 has been based on the findings from research question 1 and translated to the human-ECA situation.

We have looked at anthropomorphic characteristics of ECAs (gender, attitudes, emotion, personality, ethnicity) and what role they play in developing a friendship relation. Answering research question 3 has been done by presenting a way to create 'friendlier' ECAs that incorporates the findings of research question 2.

4. On Friendship

Interpersonal relationships are a subject of research in social psychology. Theories explaining why people start and maintain relationships are numerous. In reinforcement theory it is assumed that we are attracted to people if we are rewarded in the presence of that person. Rewards don't have to be material. Confirmation of attitudes for example also qualifies as reward. An example is Byrne's 'Law of Attraction': The attraction towards a person A is a positive linear function of the proportion of positive reinforcements (positive reinforcement divided by total reinforcements) received from A. In social exchange theory it is assumed that rewards must outweigh the costs. Commitment to a relation relies on investments, rewards and available alternatives of the relation. In equity theory it is assumed that fairness is a central issue in relationships. The perceived input/outcome ratios of both partners involved are equal. Friendship is a specific kind of interpersonal relationship. An accepted definition (cited in [9]) is:

Definition (Friendship) a voluntary interdependence between two persons over time, that is intended to facilitate social emotional goals of the participants, and may involve varying types and degrees of companionship, intimacy, affection, and mutual assistance.

Stages of friendship can be distinguished: an initiation phase (assessment of attraction, disregard criteria), a maintenance phase (conversations about attitudes, personal issues) and a termination phase due to negative changes in the relationship.

Aspects of friendship that need to be considered are gender (e.g., activity-based men's friendship vs. affectively-based women's friendship), age (commonality vs. relational), social class and ethnic background. Effects of friendship on interaction include increase of altruistic behavior, a positive impact on task performance (and attribution of task success) and an increase in self-disclosure, including showing emotions.

Interpersonal attraction is an important factor in friendship. It is governed by positive reinforcements ('Law of Attraction', see above) and similarity between subjects is a key factor: to put it simple, we don't dislike people that are just like us. Hence, similarity of attitudes, personality, ethnicity, social class, humor, etc., reinforces the relationship. Other issues are physical attractiveness (the 'halo effect') and reciprocity of liking (whether we think that the other person likes us). The perceived reciprocity of liking is especially important in the initial stage of friendship formation.

5. Translating Friendship

How do we translate the issues that play a role in human−human friendship to issues that can be implemented in human−ECA friendship? Obviously, the CASA studies play a role here. In these 'Computers Are Social Actors' studies [15] evidence is gathered to support the *Media Equation*: Media equate real life, which means that people respond to mediated worlds as if they were real. Special attention is given to the CASA paradigm. This can be viewed as a specialization of the anthropomorphic tendency of humans. By means of several user studies it is shown that people respond to computers as if they were humans. Given minimal cues, people attribute personality and gender stereotypes to computers, and respond to automated flattery as if it were given by a human. The research also shows that users do not believe the computers *are* human, nor that their complexity makes them *like* humans. The results indicate that users respond to computers as *if* they were humans, an important finding for our research. It must be noted that the cues that were used to elicit the anthropomorphic responses were minimal. Word choice elicited personality attribution, voice pitch elicited gender attribution. In ECAs, the cues are not even minimal. Gender can be communicated by means of physical appearance and voice, personality can be communicated by means of behavior, word choice and nonverbal communication, much like humans. Consequently, the CASA paradigm should be applicable to ECAs at least as well as to computers in general.

Hence, users will attribute gender, attitudes, personality, emotion, ethnicity and other human characteristics to ECAs. These issues are instrumental in friendship formation between humans and therefore translation to the human−ECA situation will make it possible to generate human−ECA friendship. Among the issues we have to look at are the possible implementation of a reinforcement theory. From that perspective, the way to create a relationship between a human and an agent is to
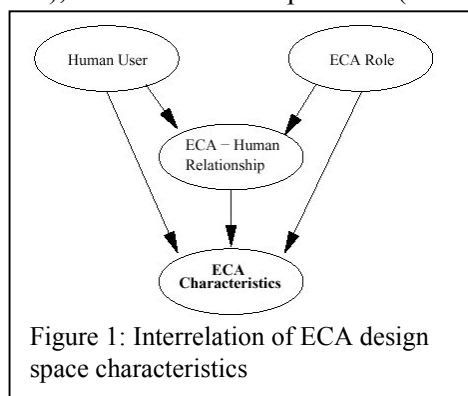
provide the human with rewards. These rewards can be numerous: fun, information, or other positive reinforcements. Hypothetically, if we use positive reinforcement, the human will become attracted to the agent. Social exchange theory dictates that the rewards of a relationship must outweigh the costs. Hypothetically, the same general principle can be applied to human−ECA relationships: rewards for the human need to be higher than costs for the human. Examples of rewards for humans are fun interactions or positive reinforcements, examples of costs are irritation or negative reinforcements. Equity theory dictates that the perceived input/outcome ratios of the parties involved should be equal. Hypothetically, this should be true for the human−ECA situation too. This means that the ECA should not be the only one to contribute towards a relationship, but the ECA should also 'expect' contributions from the user and 'gain' something out of the relationship. Communicating reciprocity of liking by an ECA could be an application of this theory.

In order to have a friendship relation with a human, the ECA needs to obey also the rules of human-friendship relations that deal with time. The friendship relationship needs to be initiated and maintained, and in each stage the ECA needs to conform to the rules that govern these stages of friendship, including a stage of decline. This requires different behavior during the different stages, as explained in the previous section.

Further considerations about the translation of mechanisms involved in friendship formation to the human−ECA situation include examining the influence of certain human characteristics. For example, when users are younger (or males) they may base the relationship with the ECA more on commonality, whereas older users (or females) may deem the relational features of a friendship as more important. Sexual attraction, social class and age similarity, and ethnicity also influence, as discussed in section 4, attraction and friendship and need to be considered.

6. The Design Phase

The most straightforward way to incorporate the factors affecting friendship is to take the user characteristics, domain role and intended interpersonal role into consideration in the design process of he ECA, and incorporate them in the decisions about the characteristics and competences of the ECA (Figure 1). These are decisions on the action and interaction level. For the characteristics we need to be aware of the stereotypes attached to the characteristics (see also section 8). The representational and algorithmic level is the level of knowledge and competence representation, the 'architecture' level. Competences include domain competences (the necessary, sufficient and desirable domain knowledge and competences an ECA should possess in order to be perceived competent in its domain role), the relational competences (social knowledge and competences an ECA should possess in order to be perceived competent in its interpersonal role) and communication competences (issues regarding the conveying of domain role and relational information). Obviously, relational competences are dependent on the intensity of the desired human−ECA relation. An example of a relational competence is the degree of user modeling that the ECA is able to do in order to 'get to know' the user. An example of communication competence (both verbal and nonverbal) related to the interpersonal role of the ECA is the ability to communicate reciprocity of liking.

Figure 1: Interrelation of ECA design space characteristics

In order to decide about the characteristics and competences a high abstraction level ECA design tool is needed, a *design method for the character of an ECA*. This design method has to facilitate consideration of the interrelation of user characteristics, ECA characteristics and roles in order to increase the likelihood of friendship formation. We propose a design process that is based on general scenario-based design and its specialization in [6] with focus on the earlier mentioned friendship aspects of ECA design.

Scenario-based ECA character development is an iterative prototyping method using scenarios, story boarding, and group techniques. In the process, the developer collaborates with the user to build a model of the ECA's character by means of scenarios, storyboards, and sample interactions, that enable

the user and the developer to envision the ECA in its future role. The aim of the character to be designed is to be tailored to the specific user group characteristics in order to increase the friendship likelihood. This is the reason why future users and social psychology experts play an important role in the design process.

There are several reasons why scenario-based design is practical as the design process of ECA characters. First, scenario-based design has already been used successfully in developing ECAs, so its validity is already shown. Furthermore, this application of scenario-based design in ECA design showed that the method is useful for putting pieces of different research in ECA technology together. Most importantly, the design method is based on designing explicitly with social interaction in mind, and can be used to focus on the 'soft' side of the ECA design, the personality, roles, and competences.

7. Adapting to the Human Partner

".. and comes face to face with a receptionist [avatar]. For a moment, he can't peg her racial background; then he realizes that this [avatar] is half-black, half-Asian – just like him. If a white man had stepped off the elevator, she probably would have been a blonde. A Nipponese businessman would have come face to face with a perky Nipponese office girl."

Snow Crash, Neal Stephenson, 1992

Another way to increase friendship likelihood is to align the characteristics of the user and the ECA during interaction in order to facilitate friendship and attraction by *adapting the character* of the ECA to the perceived characteristics of the user (see Figure 2). The *ECA characteristics updating* part of the ECA changes the ECA characteristics according to some adaptation strategy. This strategy should be directed at finding the optimal set of characteristics for friendship, given the user model. Some form of user modeling should enable the ECA to extract the user personality and other characteristics from the user's input. Because the general tendency of humans is to like people that are similar to them, a plausible strategy could focus on adopting user characteristics (personality, attitudes) that are similar to those of the user.

Not all characteristics can be adapted. Certain issues in the environment of the ECA limit the adaptable characteristics. For example, if the personality of a sales-ECA would be adapted to be completely similar to the personality of a user, problems will arise when an extremely introvert user uses the agent. Constraints like this need to be considered in the design phase of the agent, and integrated in the adaptation strategy of the ECA.

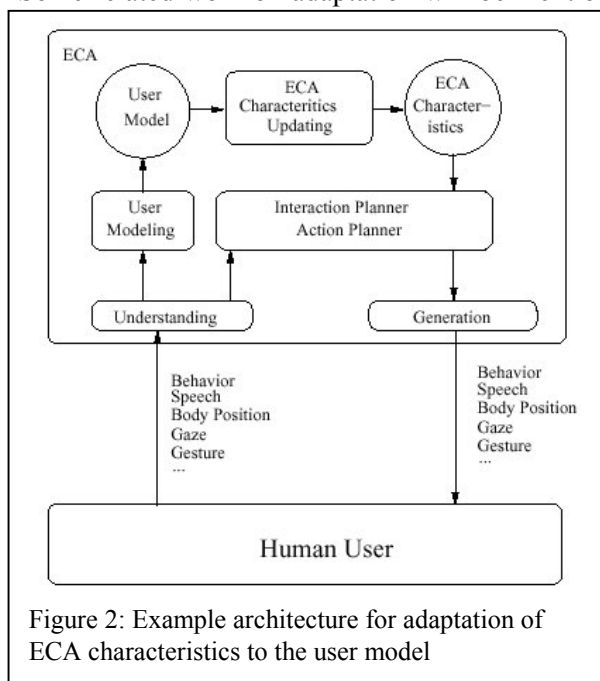Some related work on adaptation will be mentioned below. Adapting the computer's personality has been shown in user studies in a laboratory setting [15] to increase perceived liking for a computer communicating a certain 'personality' with minimal (pre-programmed) cues. It showed that when a user was an introvert, and the computer changed its personality from extravert to introvert during the interaction, the perceived liking increased. This was also true for an extravert user and a computer personality changing from introvert to extravert.

Figure 2: Example architecture for adaptation of ECA characteristics to the user model

The idea of adapting personality by some strategy is also mentioned in the 'Peedy' project [1]. In this ECA the emotional state and personality of the user are assessed using Bayesian belief networks. After assessing the state of the user, the emotional state and personality of the ECA can be changed according to some strategy. Then, theoretically, the psychological state of the ECA can be simulated, partly by using the same Bayesian belief networks.

Although in the Relational Agent project at the MIT [3] the ECA does not adapt to the user, the results from the user study indicate that adapting ECA behavior to the user personality would be effective. A study showed that extravert users liked a small-talking ECA better than a non-small-taking ECA, whereas introvert users liked a non-small-talking ECA better than a small-talking ECA. Obviously, considering small talk as a behavior that is exhibited more by extravert humans than by introvert humans, this liking is explained by our previous observations. If the personality of the user would be perceived by the ECA in one way or the other, the communicated personality of the ECA (by use of small talk) could be adapted to be similar to the personality of the user. This way, the extravert user as well as the introvert user would be optimally attracted to the ECA.

8. Stereotypes Revisited

Stereotypes are attributions of certain characteristics to certain groups. The cognitive basis of stereotypes is the fact that humans organize their environment by using schemas and categorization. Not only objects are organized ('cars', 'flowers'), but also humans ('secretaries', 'programmers'). Subconsciously, certain characteristics are attributed to these social categories. These are the stereotypes, generalizations of the world around us. An example of a stereotype is 'secretaries are women'. Stereotyping helps processing information by reducing its complexity.

Stereotyping exists in almost every aspect of our life. Examples of groups of people about whom stereotypes exist are gender, occupation, race, sexual orientation and age groups. When meeting a member of a group, humans expect certain characteristics of that member, based on the characteristics we attribute to the stereotype of that group. When humans have a meeting with a manager, most are surprised when that manager turns out to be a woman, because for most people the stereotypical manager is a man.

Stereotypes persist. Information that is consistent with the stereotype is remembered, information that is inconsistent is easily forgotten. Only after intensive personal contact with a member of a stereotyped group will the personal information prevail above the stereotype information.

The dynamics between the stereotypes of the ECA's characteristics and the domain and interpersonal role make decisions about characteristics not straightforward. The optimal choice of ECA characteristics relies on a careful consideration of the different stereotypes attached and the domain and interpersonal role. For example, we may want to create an ECA that sells cars (domain role) and tries to get real close to its user (interpersonal role), and we are going to decide on the gender. On the basis of the interpersonal role, we may want to decide the female gender for the ECA. However, a user survey may conclude that the stereotype the intended users have about females does not include 'knowledge about car mechanics'. Therefore the domain role is influenced negatively by choosing the female gender. The optimal gender can only be decided upon by careful consideration of the importance of each role.

A note on the ethics of using stereotypes is also in order. Stereotypes are used in ECA character design to increase effectiveness and elicit certain responses of the user that are based on that stereotype. In the car sales example, the male gender can be used to communicate domain competence. However, by using the stereotypes we reinforce the stereotype. Stereotypes may seriously harm people by dominating our conception of them, as described above. Hence, designers should consider the proactive stance, and design ECAs a-stereotypical in order to change the cultural determined stereotype. The disadvantage of using a-stereotypical ECAs is that designers cannot benefit from the (positive) stereotypical perceptions of the user. Ethics and need for effectiveness need to be carefully weighed [10].

9. Conclusions

For the moment we had to confine ourselves to a theoretical research perspective, neither the design method proposed nor the adaptive architecture could be applied or tested. In order to test the scenario-based character design method ECA characters should be developed using his method. Depending on the experiences of the design process, the proposed method should be further refined and adapted.

Research in adaptive ECA architectures should be continued by implementing an ECA architecture that adapts to the user, based on the technique proposed in our research.

We have introduced earlier several ECAs with modest verbal and nonverbal communication capabilities ([8], [11, 12, 13, 14]). Our research programme also includes designing agents that have personality and show emotions [7]. In this paper we made a first step towards introducing ECAs that work on the formation of long-term relationships with their human conversational partner. Implementation and evaluation are the necessary next steps.

References

[1] G. Ball and J. Breese. Emotion and personality in a conversational agent. In: [5].

[2] D. Byrne. Attitudes and attraction. *Adv. in Experimental Social Psychology*, 4: 34–89, 1969.

[3] T. Bickmore and J. Cassell. Relational agents: A model and implementation of building user trust. In *ACM CHI 2001 Conference Proceedings*, Seattle, Washington, 2001.

[4] J. M. Carroll (ed.). *Scenario Based Design: Envisioning Work and Technology in System Development*. John Wiley, New York, 1995.

[5] J. Cassell et al. (eds.) *Embodied Conversational Agents*. MIT Press, Cambridge, 2000.

[6] E.F. Churchill. May I help you? Designing embodied conversational agent allies. In: [5].

[7] Duy The Bui, D. Heylen, M. Poel & A. Nijholt. ParleE: An adaptive plan based event appraisal model of emotions. Parlevink Internal Report, University of Twente, April 2002.

[8] I. van Es, D. Heylen, B. van Dijk & A. Nijholt. Gaze behavior of talking faces makes a difference. In: Proc. *ACM-CHI 2002*, Minneapolis, April 2002, to appear.

[9] B. Fehr. *Friendship Processes*. SAGE Publications, Thousand Oaks, California, 1996.

[10] B. Friedman (ed.). *Human Values and the Design of Computer Technology*. Cambridge University Press, Stanford, 1997.

[11] D. Heylen & A. Nijholt. Designing and implementing embodied agents: Learning from experience. In: Proc. *Multimodal Communication and Context in Embodied Agents*. C. Pelachaud & I. Poggi (eds.), Workshop at 5th Intern. Conf. on *Autonomous Agents (Agents 2001)*. Montreal, 77-82. Also: *CEUR Workshop Proc.* Vol-46, 2001; http://ceur-ws.org.

[12] D. Heylen, A. Nijholt and M. Poel. Embodied agents in virtual environments: The Aveiro project. *European Symp. on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems*, Tenerife, 2001, Verlag Mainz, Aachen, 110-111.

[13] A. Nijholt. From virtual environment to virtual community. In: *New Frontiers in Artificial Intelligence.* T. Terano et al. (eds.). LNCS 2253, Springer Verlag, Tokyo, 2001, 19-26.

[14] A. Nijholt. Agents, believability and embodiment in advanced learning environments. In: Proc. IEEE *Intern. Conf. on Advanced Learning Technologies (ICALT 2001)*, T. Okamto et al. (eds.), Madison, 2001, IEEE, Los Alamitos, 457-459.

[15] B. Reeves and C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, 1996.

[16] B. Stronks. Friendship relations with embodied conversational agents: Integrating social psychology in ECA design. Report, Parlevink Research Group, University of Twente, 2002.