

A BDI Dialogue Agent for Social Support: Specification and Evaluation Method

J.M. van der Zwaan
Delft University of Technology
Jaffalaan 5
2628 BX Delft, The
Netherlands
j.m.vanderzwaan@tudelft.nl

V. Dignum
Delft University of Technology
Jaffalaan 5
2628 BX Delft, The
Netherlands
m.v.dignum@tudelft.nl

C.M. Jonker
Delft University of Technology
Mekelweg 4
2628 CD Delft, The
Netherlands
c.m.jonker@tudelft.nl

ABSTRACT

An important task for empathic agents is to provide social support, that is, to help people increase their well-being and decrease the perceived burden of their problems. The contributions of this paper are 1) the specification of speech acts for a social support dialogue agent, and 2) an evaluation method for this agent. The dialogue agent provides emotional support and practical advice to victims of cyberbullying. The conversation is structured according to the 5-phase model, a methodology for setting up online counseling for children. Before this agent can be used to support real children with real-world problems, a careful and thorough evaluation is of utmost importance. We propose an evaluation method for the social support dialogue agent based on multi-stage expert evaluation in which (adult) online bullying counselors interact with the system with varying degrees of freedom. Only when we are convinced that performance of the system is satisfactory, children will be involved, again in multiple stages and under the supervision of experts.

Categories and Subject Descriptors

I.2.1 [Artificial Intelligence]: Applications and Expert Systems

General Terms

Design, Experimentation

Keywords

Conversational agents, Verbal and non-verbal expression, Modeling cognition and socio-cultural behavior

1. INTRODUCTION

Social support refers to communicative attempts to alleviate emotional distress and is aimed at increasing the well-being of people and decreasing the perceived burden of their problems. Recent developments in affective computing show that empathic agents are increasingly capable of complex social and emotional dialogues. However, these dialogues are predominantly task oriented, i.e. to help the user perform

a concrete task, such as finding information or learning [15, 16].

Generally, giving social support is unrelated to this type of tasks; it is typically a non-task oriented effort. In our research, we are investigating how and to what extent Embodied Conversational Agents (ECAs) can provide social support. Recently, we proposed a design for an ECA that gives social support to children that are victims of cyberbullying [27]. Cyberbullying refers to bullying through electronic communication devices [17]. It is a complex problem that has a high impact on victims [18]. Research shows 40–60% of the victims is emotionally affected by incidents of cyberbullying [18, 20]. The anti-cyberbullying ECA tries to empower these victims by giving emotional support and practical advice.

The anti-cyberbullying agent implements different (verbal and non-verbal) strategies for giving social support. This paper is focused on the dialogue engine of the anti-cyberbullying agent, i.e. verbal strategies for social support. Therefore, the embodiment and non-verbal behavior of the agent are beyond the scope of this paper. In the remainder of this paper we use the term ‘dialogue agent’ to refer to the dialogue system and ‘anti-cyberbullying agent’ to refer to the complete system (the dialogue system combined with the embodiment).

Cyberbullying is a real problem, affecting real people. It is not our intention to present the anti-cyberbullying agent as a solution to cyberbullying. As mentioned before, our focus is on providing social support. Given the sensitivity of the topic and the vulnerability of the target audience (children), a careful and thorough evaluation is highly important. In this paper, we present 1) our implementation of different types of verbal social support and 2) our evaluation plan for the dialogue agent.

The paper is organized as follows. In section 2, we discuss related work on (embodied) conversational agents. In section 3 we operationalize social support for the anti-cyberbullying agent. Section 4 introduces the architecture of the dialogue agent. In section 5, we specify the social support types and explain how they were implemented in the prototype. Section 6 presents our plan for the evaluation of the dialogue system. Finally, in section 7, we present our conclusions.

2. RELATED WORK

Early work on affective computing demonstrated that agents are able to reduce negative emotions in users by addressing them [13]. Since then, emotional agents have been applied

Appears in: *Proceedings of the Workshop on Emotional and Empathic Agents, in the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, June, 4-8, 2012, Valencia, Spain.

predominantly in task oriented systems, i.e. systems that support users in performing concrete tasks, such as finding information. Examples include museum guide MAX that provides users with information about the museum and exhibitions [15], and agent GRETA that presents health information to the user [21]. Another popular application of emotional agents is responding to user emotions in e-learning and tutoring systems [8, 16, 26]. These so called pedagogical agents use different (emotional) strategies, such as displaying active listening behavior, encouragement and praise, to motivate the user and to make learning more engaging.

Cavazza et al. developed the ‘How was your day’ (HWYD) system, a non-task oriented ECA that allows users to talk about their day at the office [5, 23]. The system tries to influence the user’s attitudes as a part of a free conversation on work related topics, such as office mergers, promotion and workloads. The system alternates between employing clarification dialogue (asking questions to find out details) and generating appropriate affective responses to the information gathered. The system allows users to speak uninterrupted for longer periods of time (utterances of > 30 words). In addition to short sympathetic responses to the user’s input, the system may start a longer utterance to provide advice and support. These longer utterances are called comforting tirades. Comforting tirades are aimed at encouraging, comforting or warning the user. An important difference between the HWYD system and the anti-cyberbullying agent is the structure of the conversation. While the HWYD system incorporates social support into free conversation, the anti-cyberbullying agent imposes a structure on the conversation. This structure facilitates giving support, because the agent’s verbal support actions are linked to this structure (see section 3).

Small talk is non-task oriented talk; it is not used for content exchange, but has a social function in the conversation. Giving social support has certain similarities with small talk. For example, almost all social support categories can be found in the small talk taxonomy presented by Klüwer [14]. However, small talk is also typically used in task oriented systems, for example real-estate agent REA uses small talk to make the user feel comfortable before asking questions about sensitive topics such as money [3]. For giving social support, our dialogue agent uses a sequence similar to the one defined by Schneider for small talk [22]:

1. A query from the dominant conversation partner (in our case, this is the dialogue agent),
2. An answer to the query,
3. A response to the answer, consisting of one of the following possibilities: echo-question, check-back, acknowledgement, confirming an unexpected response, positive evaluation,
4. An unrestricted number of null steps or idling behavior.

Generally, the dialogue agent will give support in step 3 of the model, for example by responding sympathetically to the user’s answer to the query.

3. SOCIAL SUPPORT

Schneider’s model specifies the dialogue agent gives support in response to the user, but it does not show what kind

Support type	Description	Example
Sympathy	Express feelings of compassion or concern	How awful that you are being bullied!
Encouragement	Provide recipient with hope and confidence	I know you can do it!
Compliment	Positive assessments of the recipient and his or her abilities	Good of you to have told your parents!
Advice	Suggestions for coping with a problem	Perhaps you should tell your parents.
Teaching	Factual or technical information	You can block a contact by clicking the ‘block’ button

Table 1: The types of social support implemented in the conversational agent.

of social support is given. In this section, we provide a background on social support. The agent’s verbal social support actions are based on a typology of social support in online settings [4]. This typology is relevant for the dialogue agent, because online communication is mostly textual and does not depend on additional communication channels (such as non-verbal behavior and auditory information). The typology consist of five main support categories [4]:

- Information support (messages that convey instructions),
- Tangible assistance (offers to take concrete, physical action in support of the recipient),
- Network support (messages that appear to broaden the recipient’s social network),
- Esteem support (messages that validate the recipient’s self-concept, importance, competence, and rights as a person), and
- Emotional support (attempts by the sender to express empathy, support the emotional expressions of the recipient or reciprocate emotion)

Each category breaks down into multiple subtypes. From these subtypes, 5 that occurred frequently in counseling conversations by chat [10] were selected to be implemented in the dialogue agent, that is sympathy (emotional support), compliment (esteem support), encouragement (emotional support), advice (information support) and teaching (information support). Table 1 lists descriptions and examples of these support types.

To facilitate giving social support, the conversation between the user and the dialogue agent is structured according to the 5-phase model. The 5-phase model was developed as a methodology to structure counseling conversations via telephone and chat [2]. The five phases of a conversation are:

1. Warm welcome: the counselor connects with the child and invites him to explain what he wants to talk about

2. Clarify the question: the counselor asks questions to try to establish the problem of the child
3. Determine the objective of the session: the counselor and the child determine the goal of the conversation (e.g., getting tips on how to deal with bullying)
4. Work out the objective: the counselor stimulates the child to come up with a solution
5. Round up: the counselor actively rounds off the conversation

The 5-phase model thus a template for the conversation. Even though multiple conversation objectives are possible, we assume the user wants to get advice on how to deal with a cyberbullying incident. Therefore, the third conversation phase has a trivial implementation; the objective of the conversation is fixed to ‘get advice on how to deal with cyberbullying’. The 5-phase model assumes the child itself can come up with a solution. Since our goal is to demonstrate how a conversational agent can give verbal social support, we relax this responsibility and have the dialogue agent take the lead in phase 4. Additionally, to simplify the model, we assume certain types of support only occur in certain phases: sympathy, compliment and encouragement can occur in phase 2; advice and teaching only occur in phase 4.

4. ARCHITECTURE

Figure 1 shows the different components of the dialogue agent’s architecture. This architecture is based on the generic architecture for companion agents and robots by Steunebrink et al. [24]. The reasoning engine is modeled according to the Belief-Desire-Intention (BDI) paradigm [6]. This means the dialogue agent has beliefs (e.g., about what advice to give in which situations), goals (e.g., to give social support), and plans (e.g., the 5-phase model). Grey boxes indicate components of the dialogue agent that have not been implemented in the prototype, i.e. the input interpretation and utterance formulation modules, the user profile and the emotional module. These components will be added to the dialogue in the future. Components that have been implemented are discussed next.

4.1 Input/Output

The agent and the user communicate through natural language text messages. Given the complexity of interpreting and generating natural language, in the current system, text interpretation and generation have not been implemented. Instead, the input and the output of the prototype consists of FIPA-ACL communicative acts [9]. The communicative used by the dialogue agent are inform and request. Inform is used to inform the receiver that a given proposition is true. Request is used by the sender to request the receiver to perform some action, for example to perform another communicative act (i.e. to answer a question). An example of a social support communicative act is:

```
send(user, inform, compliment( incident(response,
    confronted_bully), courageous) )
```

This communicative act represents a compliment given to the user for being courageous because he confronted the bully. A translation of this communicative act to natural language could be: *I think it was very brave of you to confront the bully!*

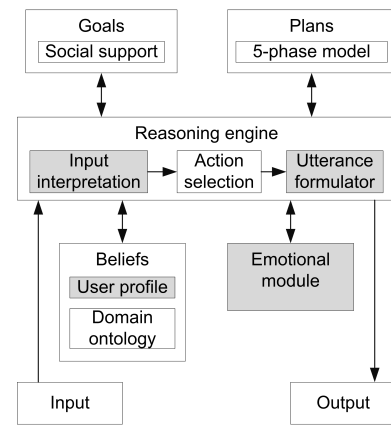


Figure 1: The architecture of the social support dialogue agent. Output is produced by the action selection engine based on the input and the agent’s beliefs.

4.2 Beliefs

The dialogue agent’s beliefs are stored in the belief base. The dialogue agent has beliefs regarding the domain (e.g., what questions to ask the user and what advice to give in different situations), social support (e.g., when to give which type of social support), and conversation management (e.g., how to open and close conversations). Additionally, the dialogue agent keeps up its beliefs about the current conversation phase, for example

```
conversation(phase, welcome)
```

and facts about the incident the buddy has learned from the user, for example

```
incident(incident_type, cyberbullying)
```

The contents of the speech acts (and thus of the conversation) are defined by the contents of the belief base. To enable reuse in other domains, the knowledge in the belief base is kept as generic as possible. This is achieved by separating dialogue management rules from domain specific knowledge. The action selection engine requests and updates information from the belief base.

4.3 Reasoning Engine

As mentioned before, the reasoning engine is based on the BDI paradigm. In the reasoning engine, beliefs are combined to select actions, which, in case of the dialogue agent, are speech acts. The main goal of the dialogue agent is to give social support. Giving social support is operationalized as completing the conversation with the user. The dialogue agent has a single plan to reach this goal, that is the 5-phase model. Beliefs about the conversation phase trigger subgoals and subsequently the dialogue agent’s actions. In phase 1 (welcome), the goal is to have greeted the user. In phase 2 (gather information), the dialogue agent has the goal of knowing certain facts about the cyberbullying incident. Established facts (i.e., the user’s answers to the dialogue agents’s questions) may trigger speech acts to give different types of social support. The implementation of social support types is explained further in section 5. The third phase of the 5-phase model (determine conversation objective) is

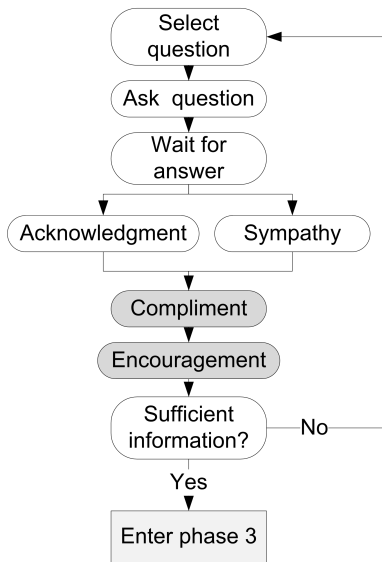


Figure 2: Social support in phase 2 (gather information). Darker grey boxes represent optional steps.

assumed to be fixed (and has a trivial implementation). In phase 4 (work out objective) the dialogue agent’s goal is to have delivered all relevant advice. The advice is based on the information the dialogue agent gathered in the second phase. Finally, in phase 5 (close conversation), the dialogue agent has the goal to have said goodbye to the user.

5. SPECIFICATION OF SOCIAL SUPPORT TYPES

Here we specify the social support types that were selected in section 3. A prototype of the social support dialogue agent was implemented in GOAL, a high level agent programming language [12]. We assume that sympathy, compliment and encouragement only occur in the second conversation phase (gather information), and advice and teaching only in phase 4 (work out objective).

After greeting the user in conversation phase 1 (welcome), the second conversation phase (gather information) starts. Figure 2 gives an overview of phase 2. Phase 2 consists of a recurring pattern of the dialogue agent selecting and asking a question, the user answering that question, and the dialogue agent acknowledging the answer. An acknowledgement is either neutral (e.g., *I see*, or *Okay*) or sympathetic. In addition to acknowledging the input (either neutrally or sympathetically), the dialogue agent optionally compliments the user or encourages him. If the dialogue agent has gathered sufficient information (what is sufficient depends on domain knowledge), it enters the third conversation phase (determine conversation objective), which, in the prototype, has a trivial implementation; the dialogue agent assumes the user wants advice on how to deal with cyberbullying. The advice is delivered in phase 4 (work out objective), which is illustrated in figure 3. After selecting a piece of advice, the dialogue agent presents it to the user. If the dialogue agent advises the user to perform a task that requires technical knowledge, it will follow up with the question whether the user wants him to explain how to perform the task. If the user confirms, the dialogue agent explains how to perform

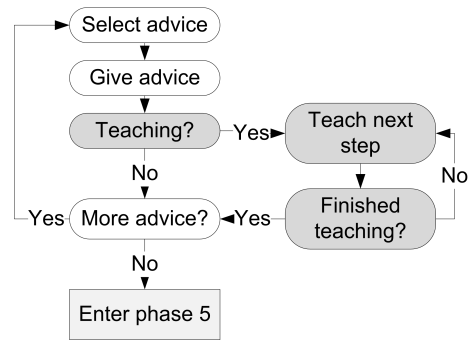


Figure 3: Social support in phase 4 (work out objective). Darker grey boxes represent optional steps.

the task step by step. If the dialogue agent has given all relevant advice, the fifth phase (round off) is entered and the dialogue agent says goodbye to the user.

5.1 Sympathy

Sympathy expresses feelings of compassion or concern. During the information gathering phase, the dialogue agent may respond sympathetically to answers given by the user. The dialogue agent expresses sympathy if it follows from his beliefs sympathy is applicable, otherwise it plays safe by staying neutral. The implementation of sympathy is illustrated in the following example:

Dialogue agent: *Can you tell me what happened?*
Child: *Someone is calling me names on msn*

The child’s utterance causes the addition of the following incident facts:

```
incident(type_cb, name_calling).
incident(method_cb, msn).
```

to the belief base of the dialogue agent. Since the belief base also contains the following fact:

```
sympathetic_acknowl(type_cb, name_calling) :-
    incident(type_cb, name_calling).
```

the agent responds sympathetically to the user:

Agent: *That’s awful!* (sympathy)

Absence of the `sympathetic_acknowl` rule would have resulted in a neutral acknowledgement of the user’s input:

Dialogue agent: *I see* (acknowledgment)

5.2 Compliment

Compliments are positive assessments of the recipient and his abilities. In the context of a social support dialogue about a specific event, there are two possibilities for the dialogue agent to give compliments: 1) the user tells the dialogue agent he performed a constructive, positive or otherwise positive action (e.g., in response to being bullies, the user didn’t retaliate), and 2) the user performs well as a dialogue partner (e.g., the user gives a clear explanation of something). Currently, only the first type of compliment

is implemented. The following example illustrates how the dialogue agent compliments the user.

Dialogue agent: *How did you respond when you were being called names on msn?*

Child: *I told him to stop, but he didn't listen*

The child's utterance causes the addition of `incident(response, confronted_bully)`.

to the beliefs of the dialogue agent. Additionally, the belief base contains the following information:

```
quality(courageous).
```

```
characteristic_of(confronted_bully, courageous).
```

```
compliment(Fact, Value, Quality):-  
    incident(Fact, Value),  
    characteristic_of(Value, Quality),  
    quality(Quality).
```

The `quality` fact states courageousness is a quality and the `characteristic_of` fact links the user response to the quality. The `compliment` rule combines the `incident` fact with the quality and the user response. This enables the agent to compliment the user:

Dialogue agent: *I see.* (acknowledgment)

Dialogue agent: *That was very brave of you!* (compliment)

In case multiple compliments are triggered by an `incident` fact, the dialogue agent randomly selects one. This procedure will be extended in future work.

5.3 Encouragement

Encouragement is about providing the recipient with hope and confidence. The process of encouraging the user closely resembles the implementation of giving compliments. Again, we assume that encouragement is always given in response to a user utterance. Utterances indicating the user's situation is severe trigger encouragement. The circumstances under which a situation can be considered severe depend on domain knowledge. The implementation of encouragement is illustrated by the following example:

Dialogue agent: *Has he bullied you before?*

Child: *Yes, all the time*

The child's response results in the addition of `incident(bullied_before, often)`.

to the beliefs of the dialogue agent. Based on the following rule in the belief base:

```
encouragement(bullied_before, often):-  
    incident(bullied_before, often).
```

encouragement is triggered and the dialogue agent encourages the user:

Dialogue agent: *I'm sorry to hear that* (sympathy)

Dialogue agent: *Let's try to stop the bullies!* (encouragement)

5.4 Advice

In phase 4, the dialogue agent gives advice on how to deal with cyberbullying. Which advice is given depends on domain knowledge and the specific situation of the user. The domain specific rules that trigger pieces of advice also provide a reason for giving the advice. The reason is added to the speech act to allow the dialogue agent to justify its advice. For example, if the belief base contains the following information:

```
incident(bully, classmate).
```

```
advice(talk_to_teacher, bully, classmate):-  
    incident(bully, classmate).
```

, the advice `talk_to_teacher` is triggered by the `incident` fact. And the dialogue agent can say something like:

Dialogue agent: *Since you are bullied by a classmate, it might be a good idea to talk to your teacher.* (advice)

Child: *Okay*

In case the same piece of advice is associated with multiple explanations, the dialogue agent randomly picks one. In future work, we will implement more sophisticated means of choosing between multiple explanations, e.g. by selecting an explanation that has been successful before. After giving a piece of advice, the dialogue agent waits for confirmation from the user. Once the user has confirmed, it moves on to the next piece of advice, or phase 5 when all advice has been given.

5.5 Teaching

Teaching concerns conveying factual or technical information. In the domain of cyberbullying, many measures the user can take to protect himself require technical know-how, such as blocking users, changing accounts, and reporting inappropriate content. The implementation of teaching is based on the work by Vergunst [25]. The dialogue agent uses a list of instructions to explain how to perform the task. After the explanation of a step, the dialogue agent waits for confirmation from the user before moving on to the next step. The following example illustrates how teaching is implemented in the prototype. The belief base of the dialogue agent contains facts describing which pieces of advice require teaching and a list of instructions to perform the teaching:

```
advice(block_msn_contact, bully, anonymous):-  
    incident(method_cb, msn),  
    incident(bully, anonymous).
```

```
teaching(block_msn_contact,  
    [login_msn, locate_contact, right_click_contact,  
    choose_block_contact]).
```

The advice rule triggers the advice to be given and the `teaching` fact states this piece of advice might require teaching and provides the list of instructions.

Dialogue agent: *Because the bully is anonymous, blocking this contact on msn might stop him.* (advice)

Dialogue agent: *Do you want me to explain to you how to do that?* (teaching)

Child: *Yes, please!*
Dialogue agent: *First, you login to msn.* (teaching)
Child: *Yes*
Dialogue agent: *Next, locate the contact you want to block in your contact list.* (teaching)
Child: *OK*
...
Dialogue agent: *Okay, now you know how to block a contact in msn.* (teaching)

If the user would have answered *No, thank you* in the second turn, the dialogue agent would have aborted the teaching and moved on to the next piece of advice.

6. EVALUATION PLAN

The dialogue system described in the previous sections is a component of an ECA that should provide emotional support and practical advice to victims of cyberbullying. In addition to the dialogue system, the embodiment of the ECA is another main component of the anti-cyberbullying agent. We propose to evaluate the anti-cyberbullying agent in three stages: first, the dialogue system and the embodiment should be evaluated separately. This can be done in parallel. After that, the complete anti-cyberbullying agent, i.e. the combination of the dialogue system and the embodiment, should be evaluated. The advantage of evaluating the separate components before the complete system is that we will be able to clearly demonstrate the contribution of individual components to the results of the final system.

This section will describe our evaluation plan for the dialogue agent (the dialogue component of the anti-cyberbullying agent). The goal of the evaluation is to determine the extent to which users experience social support when interacting with the dialogue agent. This will be measured with a questionnaire that was used by Fukkink and Hermanns in a qualitative content analysis of support provided by a Dutch child helpline [11]. Prior to interacting with the dialogue system, participants will indicate on a 9-point scale how they feel (well-being) and how severe their problem is (perceived burden of the problem). These questions will be asked again after the interaction. In addition, participants will also rate (again on a 9-point scale) to what degree they felt supported, whether they now knew what to do, if they felt they had been taken seriously, whether they had been made to feel at ease, and if they understood the dialogue system’s messages. Finally, participants will be asked to rate the trustworthiness of the dialogue agent. The perceived social support will be compared to perceived social support in conversations with human counselors.

The evaluation plan consists of multiple, incremental stages in which the dialogue system is improved based on the feedback from the previous stage before moving on to the next. If performance of the dialogue agent is unsatisfactory, the current stage will be repeated after incorporating the feedback. The different stages of the evaluation plan are listed in table 2.

6.1 Expert Evaluation

Since we are dealing with a sensitive topic (cyberbullying) and a vulnerable target audience (children), we need to know how good the system is before we involve children in the evaluation process. Therefore, we first perform an expert evaluation with adults trained to hold counseling

Participants	Experiment
Online counselors	Dialogue fragments WOZ with scenarios Dialogue system with scenarios Dialogue system with free input
Children	WOZ with scenarios Supervised dialogue system with scenarios
Cyberbullying victims	WOZ with free input Supervised dialogue system with free input

Table 2: Overview of the multi-step evaluation plan for the dialogue agent.

conversations with children about different topics, including bullying¹. These experts will be asked to interact with the dialogue agent from a children’s perspective.

Before allowing the experts to interact with the dialogue agent, they will be asked judge fragments of social support conversations. This is done to make sure the dialogue system’s messages are clear and understandable. The fragments will be similar to the example dialogues in section 5 and created from counselor utterances found in actual chat conversations. The experts will assess the fragments on understandability for children, recognizability and relevance of social support types, and the extent to which the formulation is consistent with the experience of the target audience. In addition, they can suggest alternative formulations. At the end of this stage we will have gathered a validated library with conversation fragments for the dialogue agent.

For the next stage, we will design scenarios of frequently occurring cyberbullying situations. In this stage, experts will interact with a Wizard of Oz (WOZ) system based on these scenarios. In a WOZ experiment, a human experimenter selects the utterances of the dialogue agent. Participants first read the situation description from one of the scenarios and put themselves in the shoes of the main character. Next they fill out the pre-test questionnaire, interact with the WOZ system and fill out the post-test questionnaires. Finally, the participants are asked to give feedback on how the conversation went. They will be asked to elaborate on what went well, what could be improved, and to what extent the conversation similar was to a conversation with a human counselor. Based on the feedback, the dialogue agent will be improved.

For the next experiment we follow the same procedure. However, instead of interacting with the WOZ system, participants interact with the actual dialogue agent. After processing the feedback and updating the dialogue system, participants will interact with the dialogue agent based on free input. This means the counselors can come up with situations based on their experience and ask the dialogue agent for advice.

6.2 Involving Children

If the previous experiments have been completed successfully, we can start to involve children in the evaluation pro-

¹For the development and evaluation of the anti-cyberbullying buddy we are cooperating with psychologists from the Open University (the Netherlands) and (online) counselors from Pestweb (www.pestweb.nl).

cess. All experiments in which children participate will be conducted in cooperation with and under the supervision of experts (i.e. online counselors and/or psychologists). The first stage in the evaluation with children is a WOZ experiment with the scenarios from the second experiment of the expert evaluation. Because we use scenarios, there is no need to recruit children that have experience with being cyberbullied. The wizard will be played by an online counselor. The dialogue agent will be improved based on the feedback from the experimenter.

In the next stage, children will interact with a supervised dialogue agent. This means the dialogue agent will suggest an utterance that will be sent to the participant only if the experimenter (which is again an online counselor) approved it. Additionally, if the experimenter does not approve of the suggested utterance, she can send a custom message to the participant (just as she normally does during counseling via chat). The participant will be asked to complete the questionnaires as described previously. In addition, we will take into account the number of human interventions. Finally, feedback from the experimenter is gathered: what went well and what does still need improvement? This experiment will be repeated with new participants and updated versions of the dialogue agent, until the number of interventions is acceptably low (what is acceptable will be discussed with the experts).

6.3 Involving Cyberbullying Victims

In final stage, the previous two stages are repeated, but the dialogue agent responds to actual experiences of cyberbullying victims. First, victims interact with a WOZ and after successful completion of that stage, victims interact with the supervised dialogue agent, so the experimenter can intervene at any moment. Performance is measured with the questionnaires, the number of human interventions and feedback from the experimenter.

If in this stage of the evaluation the number of human interventions is acceptably low (again, what is acceptable will be discussed with the experts) and if scores on the social support questionnaire, scores for well-being and perceived burden of the problem are close to scores obtained by human counselors the evaluation of the dialogue agent is complete. If the embodiment has been evaluated successfully, we can move on to the evaluation of the complete anti-cyberbullying agent.

7. DISCUSSION AND CONCLUSION

In this paper, we specified 5 verbal social support types: sympathy, compliment, encouragement, advice, and teaching and, inspired by a model for small talk, implemented these in a BDI dialogue agent. The dialogue agent structures the conversation according to the 5-phase model: in phase 1, the agent welcomes the user; in phase 2, the agent gathers information about the incident; phase 3 (determine objective of the conversation) has a trivial implementation in which the conversation objective is always 'get advice on how to deal with cyberbullying'; in phase 4 the agent gives advice; and in phase 5, the conversation is rounded off. Sympathy, compliment and encouragement are always given in response to user input. Advice and teaching are offered pro-actively.

Additionally, we presented an evaluation method for the dialogue agent. Because cyberbullying is a sensitive topic

and children are a vulnerable target audience, we will start with an expert evaluation and create scenarios of common cyberbullying situations for indirect evaluation. After multiple experiments and incremental improvements on the dialogue agent we intend to involve children in the evaluation process. Experiments in which children participate will be conducted always in cooperation with and under the supervision of psychologists and online counselors. Performance of the dialogue agent will be measured with questionnaires on perceived social support and trustworthiness of the agent.

Braithwaite's typology of social support contains more support types that can be implemented in the dialogue agent. In particular empathy is relevant for the anti-cyberbullying agent, because being empathic is important in supportive communication [7]. To appear empathic, the agent needs the capability to reason about emotions. Therefore an emotional module will be added to the anti-cyberbullying agent (see figure 1). We also plan to extend the dialogue agent with additional conversation techniques online counselors use to actively manage conversations, including requesting feedback (e.g., *Is that right?*), summarizing (e.g., *So, you are being bullied in school and via msn and you haven't told anybody because you are embarrassed?*), and verbalizing feelings (e.g., *You sound disappointed, are you?*).

The dialogue agent specified in this paper is part of an embodied agent. The embodiment is currently under development and will allow the anti-cyberbullying agent to give non-verbal feedback in addition to verbal feedback. The non-verbal channel will be mainly used for the expression of (empathic) emotions. Related work on empathic agents shows that text-only agents are outperformed by embodied agents [1, 13, 19]. Therefore, we expect the perceived social support will increase when a virtual character displaying appropriate emotional expressions is added to the system's interface.

The anti-cyberbullying agent is an application that addresses a real world problem. We would like to emphasize that a lot more than satisfactory performance in laboratory experiments is needed before the application can be introduced into a real world setting. Many additional criteria play a part in the feasibility and acceptability of software applications, such as the protection of privacy and other ethical and legal issues. At the very least the anti-cyberbullying agent should be able to detect and deal with cases it can not handle, either by referring the user to a specialized helpline, or call in a human counselor that takes over the conversation.

8. ACKNOWLEDGEMENTS

This work is funded by NWO under the Responsible Innovation (RI) program via the project 'Empowering and Protecting Children and Adolescents Against Cyberbullying'.

9. REFERENCES

- [1] D.C. Berry, L.T. Butler, and F. de Rosis. Evaluating a realistic agent in an advice-giving task. *International Journal of Human-Computer Studies*, 63(3):304-327, 2005.
- [2] A. de Beyn. *In gesprek met kinderen: de methodiek van de kindertelefoon*. SWP, 2003.
- [3] T. Bickmore and J. Cassell. 'how about this weather?' social dialog with embodied conversational agents. In

- Proceedings of the American Association for Artificial Intelligence (AAAI) Fall Symposium on "Narrative Intelligence"*, pages 4–8, 2000.
- [4] D.O. Braithwaite, V.R. Waldron, and J. Finn. Communication of social support in computer-mediated groups for people with disabilities. *Health Communication*, 11(2):123–151, 1999.
 - [5] M. Cavazza, C. Smith, D. Charlton, N. Crook, J. Boye, S. Pulman, K. Moilanen, D. Pizzi, R. de la Camara, and M. Turunen. Persuasive dialogue based on a narrative theory: An eca implementation. In T. Ploug, P. Hasle, and H. Oinas-Kukkonen, editors, *Persuasive Technology*, volume 6137 of *Lecture Notes in Computer Science*, pages 250–261. Springer Berlin / Heidelberg, 2010.
 - [6] P.R. Cohen and H.J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2-3):213 – 261, 1990.
 - [7] H. Cowie and P. Wallace. *Peer Support in Action: From Bystanding to Standing By*. Sage Publications Ltd, 2000.
 - [8] S. D’Mello, B. Lehman, J. Sullins, R. Daigle, R. Combs, K. Vogt, L. Perkins, and A. Graesser. *Intelligent Tutoring Systems*, volume 6094 of *Lecture Notes in Computer Science*, chapter A Time for Emoting: When Affect-Sensitivity Is and Isn’t Effective at Promoting Deep Learning, pages 245–254. Springer Berlin / Heidelberg, 2010.
 - [9] Foundation for Intelligent Physical Agents. Fipa communicative act library specification. <http://www.fipa.org/specs/fipa00037/SC00037J.html>, 2002.
 - [10] R. Fukkink. Peer counseling in an online chat service: A content analysis of social support. *Cyberpsychology, Behavior, and Social Networking*, 14(4):247–251, 2011.
 - [11] R. Fukkink and J. Hermanns. Counseling children at a helpline: chatting or calling? *Journal of Community Psychology*, 37(8):939–948, 2009.
 - [12] K.V. Hindriks. Programming Rational Agents in GOAL. In A. El Fallah Seghrouchni, J. Dix, M. Dastani, and R.H. Bordini, editors, *Multi-Agent Programming: Languages, Tools and Applications*, volume 2, pages 119–157. Springer US, 2009.
 - [13] K. Hone. Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interact. Comput.*, 18(2):227–245, 2006.
 - [14] T. Klüwer. “I Like Your Shirt” - Dialogue Acts for Enabling Social Talk in Conversational Agents. In H. Vilhjálmsson, S. Kopp, S. Marsella, and K. Thórisson, editors, *Intelligent Virtual Agents*, volume 6895 of *Lecture Notes in Computer Science*, pages 14–27. Springer Berlin / Heidelberg, 2011.
 - [15] S. Kopp, L. Gesellensetter, N.C. Krämer, and I. Wachsmuth. *Intelligent Virtual Agents*, chapter A Conversational Agent as Museum Guide – Design and Evaluation of a Real-World Application, pages 329–343. 2005.
 - [16] T.-Y. Lee, C.-W. Chang, and G.-D. Chen. Building an interactive caring agent for students in computer-based learning environments. In *Advanced Learning Technologies, 2007. ICALT 2007. Seventh IEEE International Conference on*, pages 300–304, 2007.
 - [17] Q. Li. New bottle but old wine: A research of cyberbullying in schools. *Computers in Human Behavior*, 23(4):1777–1791, 2007.
 - [18] S. Livingstone, L. Haddon, A. Görzig, and K. Ólafsson. Risks and safety on the internet: The perspective of european children. initial findings. <http://www2.lse.ac.uk/media@lse/research/EUKidsOnline/EUKidsII%20%282009-11%29/home.aspx>, 2010.
 - [19] R. Looije, M.A. Neerincx, and V. de Lange. Children’s responses and opinion on three bots that motivate, educate and play. *Journal of Physical Agents*, 2(2):13, 2008.
 - [20] J.W. Patchin and S. Hinduja. Bullies move beyond the schoolyard: A preliminary look at cyberbullying. *Youth Violence and Juvenile Justice*, 4(2):148–169, 2006.
 - [21] C. Pelachaud, V. Carofiglio, B. De Carolis, F. de Rosis, and I. Poggi. Embodied contextual agent in information delivering application. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 2, AAMAS ’02*, pages 758–765, New York, NY, USA, 2002. Acm.
 - [22] K.P. Schneider. *Small Talk: Analyzing Phatic Discourse*. Marburg: Hitzeroth, 1988.
 - [23] C. Smith, N. Crook, J. Boye, D. Charlton, S. Dobnik, D. Pizzi, M. Cavazza, S. Pulman, R. de la Camara, and M. Turunen. Interaction strategies for an affective conversational agent. In Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud, and Alla Safonova, editors, *Intelligent Virtual Agents*, volume 6356 of *Lecture Notes in Computer Science*, pages 301–314. Springer Berlin / Heidelberg, 2010.
 - [24] B.R. Steunebrink, N.L. Vergunst, C.P. Mol, F.P.M. Dignum, M.M. Dastani, and J.-J.C. Meyer. A generic architecture for a companion robot. In J. Filipe, J.A. Cetto, and J.-L. Ferrier, editors, *Proc. 5th Int. Conf. on Informatics in Control, Automation and Robotics (ICINCO’08)*, pages 315–321, 2008.
 - [25] N.L. Vergunst. *BDI-based Generation of Robust Task-Oriented Dialogues*. PhD thesis, Utrecht University, 2011.
 - [26] K. Zakharov, A. Mitrovic, and L. Johnston. *Intelligent Tutoring Systems*, volume 5091 of *Lecture Notes in Computer Science*, chapter Towards Emotionally-Intelligent Pedagogical Agents, pages 19–28. Springer Berlin / Heidelberg, 2008.
 - [27] J.M. van der Zwaan, V. Dignum, and C.M. Jonker. Simulating peer support for victims of cyberbullying. In *Proceedings of the 22st Benelux Conference on Artificial Intelligence (BNAIC 2010)*, 2010.