Improving Social Awareness Through Thought Bubbles and Flashbacks of Virtual Characters

Jeroen Linssen[1(✉)], Mariët Theune[1], Thomas de Groot[2], and Dirk Heylen[1]

[1] Human Media Interaction, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands
{j.m.linssen,m.theune,d.k.j.heylen}@utwente.nl
[2] T-Xchange, PO Box 217, 7500 AE Enschede, The Netherlands
t.f.degroot@txchange.nl

Abstract. We present two prototypes of a serious game which is a aimed at raising police officers' awareness of social stance during street interventions by letting them interact with virtual characters. We discuss the design, implementation and evaluation of a method of feedback on the police officers' game actions. This method uses thought bubbles to show the cognitive state of virtual characters, using a theory of interpersonal stances. We use thought bubbles (1) to provide direct feedback by showing the agent's current attitude, and (2) to provide delayed feedback at the start of a new scenario by showing a flashback to the previous scenario, expressing the character's overall attitude towards the player. We conducted two experiments with students from the Dutch Police Academy and found that our implementations of these forms of feedback did not lead to directly measurable learning gains.

Keywords: Serious games · Social awareness · Virtual agents · Meta-techniques · Thought bubbles · Flashback · Law enforcement

1 Introduction

Attaining and maintaining good social skills is of importance to many professions, including that of police officers. There are limited means to train social skills through role plays with professional actors. Therefore, we propose to augment the police officers' curriculum with serious games [5]. In these games, police officers train with virtual characters to experience difficult social interactions.

Providing feedback about formative assessment is a current challenge that must be tackled to improve serious games for training interpersonal skills [7]. In this paper, we present our work on LOITER (LOItering Teenagers, an Emergent Role-play), a series of prototype serious games in which police officers have to deal with loitering juveniles. In the LOITER games, we provide feedback to players during the game through *meta-techniques*. These are out-of-character techniques used in live action role play to communicate information such as thoughts of characters in the play. We investigate an implementation of this meta-technique in the form of *thought bubbles* that explain the behaviour of virtual characters.

W.-P. Brinkman et al. (Eds.): IVA 2015, LNAI 9238, pp. 250–259, 2015.
DOI: 10.1007/978-3-319-21996-7_25

The remainder of the paper is structured as follows. In Sect. 2, we discuss related research on serious games for interpersonal skills training, thought bubbles and flashbacks. In Sect. 3, we describe the model for social interaction we adopted in our serious game, and the learning goals we focus on. In Sect. 4, we address the design of LOITER-TB and the implementation of thought bubbles. We describe the experiment we conducted with this prototype in Sect. 5 and discuss the improved prototype which features flashbacks in Sect. 6. Finally, we present the findings from our experiment with LOITER-FB in Sect. 7 and wrap up the paper with conclusions and future work in Sect. 8.

2 Related Work

Several approaches to feedback and in-game assessment to stimulate learning about social interactions have been studied. *BiLAT* (Bi-Lateral Negotiation) is a system for negotiation training with virtual characters [2]. Having played through a scenario, students discuss the events with a virtual tutor in an after-action review. A similar training environment featuring inter-cultural negotiation scenarios is offered by SASO-ST [10]. SASO-ST provides users with debriefs of played scenarios, giving rudimentary feedback on the character's cognitive state.

Moments for proper feedback and reflection improve learning in game-based learning approaches [11]. Rather than providing this feedback after finishing a game, as in *BiLAT* or SASO-ST, we give such feedback during gameplay. This way, students can try to adapt their approach during the game if necessary.

Recent research has looked at the use of thought bubbles to help people with autism spectrum disorder cope with everyday interactions [4,6], showing that thought bubbles helped people to recognise the emotions of others. In our research, we use thought bubbles to enrich interactions with virtual characters by providing extra information about the cognitive state of these characters.

To help medical students practice their interpersonal skills, Cordar et al. created virtual characters suffering from depression [1]. They found that students who were shown back stories of characters showed more empathy towards both the virtual characters and real-life actors with depression. Instead of providing flashbacks that show events that occurred before the interaction between players and virtual characters, as in [1], we investigate flashbacks that reveal how the players' own actions influenced the cognitive states of the virtual characters.

3 Interpersonal Skills Training

For our serious games, we address skills that should be attained by students of the Dutch Police Academy (DPA), focusing on interpersonal skills that are of use during street interventions. Instructors at the DPA uses *Leary's Rose*, a theory of interpersonal attitudes, to teach students about social interaction. Leary's Rose is used to classify interpersonal behaviour as *stances*: combinations of varying degrees of dominance and affect, see Fig. 1a. For example, a dominant and affectionate stance can be classified as 'helpful'. A police officer may say, in a confident and friendly tone: "Surely we can work on a solution together!" If he would assume an 'aloof' stance, he might say, averting his gaze and speaking with a low voice: "Of course I don't expect your cooperation..."

In our serious game, we use Leary's Rose to model the actions of the players and virtual characters. The virtual characters' reactions are based on Leary's claim that stances 'invite' complementing behaviour: people tend to mirror the other person's affect, but act inversely on the dominant axis, see Fig. 1b.

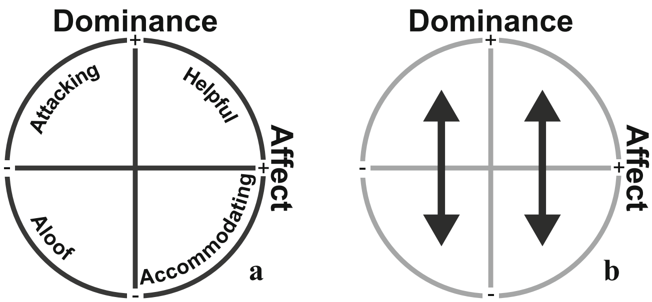

Fig. 1. (a) Leary's Rose: four interpersonal stances are as combinations of different degrees of dominance and affect. (b) 'Inviting' relationships between the stances.

In both LOITER-TB and LOITER-FB, we focus on the following learning goals:

1. Police trainees should know which stance to assume to invite a person to assume a desired stance.
2. Police trainees should be able to recognise and classify behaviour in terms of stances.

4 Thought Bubbles in LOITER-TB

In LOITER-TB, we use thought bubbles to help players achieve the learning goals from Sect. 3 by providing short-term feedback. The Serious Gaming Lemniscate Model (SGLM) asserts that players need to be drawn out of the game world and reflect on a meta-level to learn effectively [3]. For LOITER-TB, we chose to include short moments for explicit reflection.

Some live action role plays (*larps*) include *meta-techniques* offering information to players that would normally not be available to them in-character.[1] The meta-technique on which we base our use of thought bubbles is that of the inner monologue. Players can use this technique to narrate their thoughts so that other players may become aware of them. In LOITER-TB, we provide players with extra information in thought bubbles about the cognitive state of the virtual characters they interact with to help them attain the learning goals.

[1] In larps, people role play by physically acting out their characters' actions.

LOITER-TB features a short scenario that revolves around loitering juveniles playing loud music. The player's goal is to end this nuisance. He or she can do so by choosing from 4 different options on how to respond both verbally and non-verbally to the leader of the juveniles, based on the four stances from Leary's Rose. The player's choices influence how the juveniles will leave, based on the 'inviting' relations from Leary's model. For example, the juveniles may do so willingly or grudgingly. The postures of the virtual characters are based on previous research on non-verbal expression of interpersonal stances [8]. Figure 2a shows the interface of the game with a player action, the juvenile's reaction and his thoughts about the player's stance.

Fig. 2. (a) A screenshot of LOITER-TB when the player has just made a choice. (b) The visualisations of the four stances in the thought bubbles.

For the content of the thought bubbles, we used Dutch adjectives that correspond to stances in Leary's Rose [9]. Based on interviews with instructors from the DPA, we chose to include mnemonic pictures of animals that are archetypical for the different stances, as they are also used in this way in the current training curriculum of police officers, see Fig. 2b. On finishing the game, players receive a short debriefing of what happened in the scenario in which they are shown the result of their actions. The debriefing screen also indicates the stances both the player and the leader of the juveniles assumed most frequently.

5 Experiment 1: LOITER-TB

To evaluate our implementation of thought bubbles, we conducted an experiment with police trainees. Because meta-techniques provide insight in a character's thoughts, we expect that the information presented in thought bubbles aids trainees in learning about social interaction, as compared to a version of the game without the thought bubbles. We address the following three hypotheses.

Hypothesis 1. Thought bubbles help trainees to attain knowledge about social interaction that is similar to that of domain experts.

Hypothesis 2. Thought bubbles help trainees to determine which kinds of behaviour their own behaviour invites.

Hypothesis 3. LOITER-TB helps trainees to recognise stances of other people.

5.1 Method

We conducted an experiment with a repeated measures, between-subjects design with two conditions. Participants in condition TB played LOITER-TB with thought bubbles. Participants in the control condition (C) played the game without thought bubbles. The game featured eight choice points at which players chose between behaviours corresponding to the four stances (see Sect. 3).

To determine the participants' knowledge about social interaction in street intervention, we first conducted a Pathfinder network (PFnet) analysis that calculates a person's mental model based on ratings of the relatedness of given concepts [12]. By analysing the coherence of the participants' PFnets and their similarity to those of domain experts, we determined the participants' knowledge about the given concepts (hypothesis 1). Participants rated the relatedness between 10 different concepts[2] that are important to the domain of street intervention, which we constructed together with police instructors. Participants rated all combinations of concept pairs on a 7-point Likert scale, running from 'completely unrelated' to 'extremely related'.

Additionally, participants took a situational judgement test (SJT) in which they indicated which stance they would expect a juvenile to have in reaction to a given sentence. They also classified given utterance of a juvenile in terms of stance. We used these SJTs to assess participants' abilities to determine invited behaviour and to recognise stances of others (hypothesis 2 and 3).

Next, participants were instructed about the scenario and game mechanics of LOITER-TB. They played the game twice, allowing them to become more familiar with the scenario and the gameplay, and to experiment with different behaviours. Having finished the game, participants retook the PFnet test and the SJTs. They also rated their experience of the game on a number of 7-point Likert scales.

We conducted the experiment with a class of third-year police students of the DPA, who were being trained to become senior police officers ($n = 21$). These students already had practical experience by accompanying graduated police officers on the streets. The mean age of the participants was 33.0 ($SD = 6.9$); 71 % was male and 29 % female. On average, it took participants 30 min to complete the experiment, of which 6 were spent on playing the game. From the 21 participants, 12 were included in condition TB and 9 in condition C.[3]

[2] Attacking, aloof, accommodating, helpful, de-escalation, escalation, giving freedom, giving respect, working against each other, and working together.

[3] The inequality of the number of participants in the two conditions was caused by two participants not being able to complete the experiment due to technical issues.

5.2 Results

We compared the participants' pre- and post-test PFnets to an average of the PFnets of four domain experts (one police instructor and three researchers with expertise on Leary's Rose). Using ANCOVAs (dependent variables: post-test coherence and similarity scores, fixed factor: condition, covariates: pre-test coherence and similarity scores), we did not find significant differences between conditions for the coherence of the PFnets and for the similarity between participants' and domain experts' PFnets ($F(20) < 1, p > .05$). A paired-samples t-test showed that the coherence of participants' PFnets in both conditions increased significantly ($t(20) = 3.329, p = .003$; pre-game: $M = .370, SD = .363$; post-game: $M = .584, SD = .338$). Another paired-samples t-test did not show a significant difference ($t(20) < 1, p > .05$) between the pre- and post-game similarity between participants' and domain experts' PFnets across conditions.

An ANCOVA (dependent variable: post-game test score, fixed factor: condition, covariate: pre-game test score) did not show a significant difference ($F(20) < 1, p > .05$) between conditions for the participants' ability to indicate invited reactions of a juvenile (TB: $M = 2.420, SD = 1.084$; C: $M = 1.67, SD = .866$).

A paired-samples t-test showed no significant change ($t(20) < 1, p > .05$) in the participants' ability to recognise the juvenile's stances before ($M = 4.81, SD = .981$; correctly answered questions, out of 8) and after ($M = 5.05, SD = 1.284$) the game was played. An ANCOVA (dependent variable: post-game test score, fixed factor: condition, covariate: pre-game test score) also failed to show a significant difference ($F(20) < 1, p > .05$) between conditions for the participants' ability to recognise a juvenile's stance.

5.3 Discussion

Based on the PFnet analysis, we cannot prove hypothesis 1. We conclude that LOITER-TB aids participants in constructing more coherent mental models, but not in constructing mental models that are more similar to those of domain experts. The addition of thought bubbles in LOITER-TB does not improve the construction of mental models. Because the participants were already experienced at street interventions, we believe that they may already have had a good ability of analysing social interactions and did not get much added value from the feedback in the thought bubbles. Alternatively, it may have been the case that the provided feedback was not explicit enough. The SGLM suggests that players need to be drawn out of the game world to reflect on a meta-level. Our implementation of thought bubbles in the game may not have achieved this effect, which calls for closer scrutiny of how effective feedback should be provided.

Based on the analysis of the SJT, we cannot prove hypothesis 2 and 3. Because the police officers were experienced at street interventions, they already were adept at recognising stances. Therefore, there may be a ceiling effect, as they also indicated that they did not learn very much ($M = 3.24, SD = 1.57$ on a 7-point Likert scale). We expect that the learning effect of LOITER-TB may be stronger for less experienced police officers or trainees.

Participants commented on the difficulty of rating the word pairs for the PFnet analysis. Combined with the fact that we were not able to measure differences between the two conditions, we conclude that PFnet analyses may not be entirely suitable to determine people's mental models about social awareness.

6 Flashbacks in LOITER-FB

The main new meta-technique of LOITER-FB is the addition of flashbacks. These flashbacks refer to events that took place earlier in the game, thus providing feedback on players' behaviour during the game. We implemented flashbacks as thoughts of the juvenile and presented them in thought bubbles, see Fig. 3a. The flashback shows a past action and explains how the player's past average stance influences the juvenile's current preferred stance. Thus, the flashback provides both a retrospect on what happened and a prospect to what may happen.

We wrote two additional scenarios for LOITER-FB: one about the juveniles bothering shopping public, and one about the juveniles loitering at a business park. Thus, players interact with the group three times and have to bond with the juveniles. We modified the cognitive model of the virtual juvenile so that it takes the player's behaviour during previous scenarios in the same game into account. Internally, the player's choices correspond to stances in a two-dimensional grid resembling Leary's Rose. The average of these choices is then calculated as a point somewhere in that grid. Based on this average, the juvenile will have a preferred stance, namely the one that is invited by this average (see Fig. 1b). In the second and third scenario, he bases his follow-up action not only on the stance that the player's current action invites, but also on his preferred stance.

Based on comments from participants and police instructors who played LOITER-TB, we created an alternative visualisation of Leary's Rose in the thought bubbles of the virtual juvenile, see Fig. 3b. Instead of showing pictures animals, the thought bubbles show Leary's Rose with an emphasis on the stance that the player assumed. To prevent information overload, we show the thought bubble and the juvenile's reaction in sequence, and not at the same time.

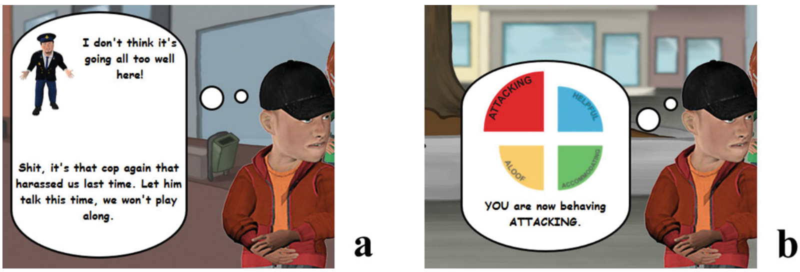

Fig. 3. (a) A thought bubble that shows a flashback. (b) The improved visualisation of Leary's Rose in the thought bubbles.

7 Experiment 2: LOITER-FB

We conducted an experiment with LOITER-FB similar to that in Sect. 5.

7.1 Method

For this experiment, we again used a repeated-measures, between-subjects design with two conditions. One group played the game with thought bubbles and flashbacks (condition FB) and the other played the game without (condition C). In total, participants made 24 choices (in 3 scenarios) when interacting with the juvenile. In condition FB, 12 thought bubbles of the virtual juvenile were shown as well as two flashbacks at the beginning of the second and third scenarios.

Instead of PFnet analyses, participants took an improved SJT. We provided more context in the situation descriptions and let participants rate the effectiveness of four given approaches (based on the four stances) to making the juveniles aware of their behaviour. This involved 8 items that were rated on a Likert scale from 1 to 7 (SJT1). In the second SJT, participants rated four stances of a juvenile on the likelihood of being invited by an utterance of the police officer, for a total of 16 items (SJT2). We compared the responses of participants to those of domain experts (two police instructors), rather than relying on preconceived notions about how these situations should be judged. We constructed similarity scores based on the differences between the ratings of the participants and the average of the ratings of the domain experts. At the end of the experiment, participants filled in the same self-report questionnaire as in Experiment 1.

In total, 28 senior police trainees completed the experiment. Their mean age was 25.9 ($SD = 7.3$); 72 % was male and 28 % female. Participants completed the experiment in 30 min; 15 played condition C, 13 condition FB.

7.2 Results

Table 1 shows the means of the participants' similarity scores on both SJTs and their increase in scores.[4] Independent samples t-tests did not show a difference between conditions on initial scores ($t(24) < 1, p > .05$), nor on post-game scores ($t(24) < 1, p > .05$) on SJT1. A one-sample t-test ($t(25) < 1, p > .05$) showed that the change in the scores of all participants did not significantly differ from 0. An ANCOVA (independent variable: post-SJT1 score; fixed factor: condition; covariate: pre-SJT1 score) showed no significant difference between condition C and FB ($F(25) < 1, p > .05$).

Independent samples t-tests did not show a difference between conditions on initial scores ($t(24) = -1.794, p > .05$), nor on post-game scores ($t(24) < 1$, $p > .05$) in SJT2. A one-sample t-test ($t(26) = 1.942, p > .05$) showed that the change in the scores of all participants did not significantly differ from 0. An ANCOVA (independent variable: post-SJT2 score; fixed factor: condition; covariate: pre-SJT2 score) showed no significant difference between condition C

[4] We assume linearity of the Likert scale items to average data across participants.

Table 1. Participants' scores on the situational judgement tests, per condition.

Condition	SJT1 score, 8 items (M)			SJT2 score, 16 items (M)		
	pre (SD)	post (SD)	incr. (SD)	pre (SD)	post (SD)	incr. (SD)
C	9.4 (2.3)	8.9 (2.7)	−.5 (3.1)	19.3 (4.9)	22.4 (4.4)	3.1 (4.6)
FB	9.3 (4.3)	9.7 (2.8)	.3 (4.7)	23.1 (5.9)	23.2 (6.2)	.1 (4.3)

and FB ($F(27) < 1, p > .05$). In both conditions, participants were satisfied with how they solved the game's scenario ($M = 5.5, SD = .8$), but were not convinced that the game improved their understanding of how their behaviour influenced others ($M = 3.8, SD = 1.6$).

7.3 Discussion

We did not find significant changes in the participants' social awareness, neither across, nor between the conditions. Our situational judgement tests showed that, before playing the game, participants already deviated only slightly from the average of the domain experts' ratings on both SJTs. This indicates that the participants may not be able to improve their social awareness much more, as our results confirm. Because of this ceiling effect, the added value of LOITER-FB may not be apparent for this user group. Some participants commented that they believed the game presents realistic cases, but that the difference between the available actions was very clear. This made it easy for them to tell which response would be the preferred one. Future prototypes may therefore benefit from a game mechanic that would allow for less restricted input.

8 Conclusions and Future Work

We designed and evaluated two iterations of a prototype serious game for improving social awareness of police trainees. Our games feature short interactions in which players resolve conflicts with virtual characters. The central contribution of our games is the inclusion of meta-techniques that provide extra information about the interaction. We used thought bubbles to show players how a virtual character interprets their behaviour and how this influences its actions. In our evaluations, we investigated the learning gains of police trainees of the Dutch Police Academy in versions of the games either with or without meta-techniques. We did not find evidence supporting our hypothesis that our implemented meta-techniques would improve the trainees' social awareness. This may be the case because these players are more adept at (unconsciously) translating such situations into knowledge without explicitly reflecting.

For future work, we will conduct experiments with less experienced police trainees to determine the effects on learning for this user group. We hypothesise that less experienced students may profit more from constructive feedback before analysing social interactions becomes natural to them. The game mechanics of the prototypes also deserve attention: input that is less restricted than through multiple-choice decision points will make the game more challenging and may cause players to behave differently. Additionally, we would like to investigate the adaptation of the virtual characters' behaviour to the players' assessed skills, making the scenarios more or less difficult when necessary.

Acknowledgements. We thank the police instructors and students who helped with our experiments. This publication was supported by the Dutch national program COMMIT.

References

1. Cordar, A., Borish, M., Foster, A., Lok, B.: Building virtual humans with back stories: training interpersonal communication skills in medical students. In: Bickmore, T., Marsella, S., Sidner, C. (eds.) IVA 2014. LNCS, vol. 8637, pp. 144–153. Springer, Heidelberg (2014)
2. Kim, J.M., Hill Jr., R.W., Durlach, P.J., Lane, H.C., Forbell, E., Core, M., Marsella, S., Pynadath, D., Hart, J.: BiLAT: a game-based environment for practicing negotiation in a cultural context. Int. J. Artif. Intell. Educ. **19**, 289–308 (2009)
3. Koops, M., Hoevenaar, M.: Conceptual change during a serious game: using a lemniscate model to compare strategies in a physics game. Simul. Gaming **44**(4), 544–561 (2012)
4. Laffey, J., Schmidt, M., Galyen, K., Stichter, J.: Smart 3D collaborative virtual learning environments: a preliminary framework. J. Ambient Intell. **4**(1), 49–66 (2012)
5. Linssen, J.M., Theune, M.: Meta-techniques for a social awareness learning game. In: Proceedings of ECGBL 2014, pp. 697–704 (2014)
6. Moore, D., McGrath, P., Powell, N.J.: Collaborative virtual environment technology for people with autism. Focus Autism Other Dev. Disabil. **20**(4), 231–243 (2005)
7. Pereira, G., Brisson, A., Prada, R., Paiva, A., Bellotti, F., Kravcik, M., Klamma, R.: Serious games for personal and social learning & ethics: status and trends. Procedia Comput. Sci. **15**, 53–65 (2012)
8. Ravenet, B., Ochs, M., Pelachaud, C.: From a user-created corpus of virtual agent's non-verbal behavior to a computational model of interpersonal attitudes. In: Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (eds.) IVA 2013. LNCS, vol. 8108, pp. 263–274. Springer, Heidelberg (2013)
9. Rouckhout, D., Schacht, R.: Ontwikkeling van een Nederlandstalig interpersoonlijk circumplex. Diagnostiekwijzer **3**, 96–118 (2000)
10. Traum, D.R., Swartout, W.R., Marsella, S.C., Gratch, J.: Fight, flight, or negotiate: believable strategies for conversing under crisis. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 52–64. Springer, Heidelberg (2005)
11. Wouters, P., van Oostendorp, H.: A meta-analytic review of the role of instructional support in game-based learning. Comput. Educ. **60**(1), 412–425 (2013)
12. Wouters, P., van der Spek, E.D., van Oostendorp, H.: Measuring learning in serious games: a case study with structural assessment. Educ. Technol. Res. Dev. **59**(6), 741–763 (2011)