Appears in 2nd International Conference on Affective Computing and Intelligent Interaction, Lisbon Portugal, 2007

Explanatory Style for Socially Interactive Agents*

Sejin Oh[1], Jonathan Gratch[2], and Woontack Woo[1]

[1]GIST U-VR Lab.
Gwangju, 500-712, S.Korea
{sejinoh, wwoo}@gist.ac.kr

[2] Institute for Creative Technolgies, Univiersity of Southern California
13274 Fiji Way, Marina del Rey, CA 90292, U.S.A.
gratch@ict.usc.edu

Abstract. Recent years have seen an explosion of interest in computational models of socio-emotional processes, both as a mean to deepen understanding of human behavior and as a mechanism to drive a variety of training and entertainment applications. In contrast with work on emotion, where research groups have developed detailed models of emotional processes, models of personality have emphasized shallow surface behavior. Here, we build on computational appraisal models of emotion to better characterize dispositional differences in how people come to understand social situations. Known as *explanatory style*, this dispositional factor plays a key role in social interactions and certain socio-emotional disorders, such as depression. Building on appraisal and attribution theories, we model key conceptual variables underlying the explanatory style, and enable agents to exhibit different explanatory tendencies according to their personalities. We describe an interactive virtual environment that uses the model to allow participants to explore individual differences in the explanation of social events, with the goal of encouraging the development of perspective taking and emotion-regulatory skills.

1 Introduction

Imagine you have two friends that just lost their jobs at the same company. Although the company gave no explanation, Robert attributes the firing to the incompetence of his manager, and quickly applies for other positions. Jim becomes convinced his performance was inadequate. He becomes paralyzed, wondering about where he failed and sinks into depression. You've probably experiences a similar situation: the same event explained in very different ways with noticeable consequences for each individual's emotional ability to cope. In social psychology, these individuals are said to differ in their *explanatory styles*, or how they explain good or bad consequences to themselves [1]. Explanatory styles are associated with certain personality differences.

* This research was supported in part by the UCN Project, the MIC 21C Frontier R&D Program in Korea, and in part by the U.S. Army Research, Development, and Engineering Command (RDECOM).

For example, pessimists like Jim tend to internalize failure and externalize success. Pessimistic style, carried to the extreme, can be maladaptive and have negative consequences for socio-emotional development and physical health. Explanatory style can also be changed through cognitive behavioral therapy, a standard psychoanalytic technique that treats depression by teaching patients to alter their habitual ways of explaining social events. In our research, we consider how to model differences in explanatory styles, both to concretize psychological theories of emotional disorders such as depression, and to inform the behavior of interactive applications that can allow users to explore explanatory differences and encourage the development of perspective taking and emotion-regulatory skills.

In contrast with scientific explanations of physical events, people's explanations of social situations are particularly susceptible to multiple interpretations. Social explanations involve judgments not only of causality but epistemic factors such as intent, foreknowledge, free will and mitigating circumstances. For example, when being hit from behind by another vehicle, one driver might assume it was a simple accident, whereas another might assume it was a malicious intentional act and responded with rage. Faithfully modeling explanatory style requires a system that can produce such social explanations and bias them systematically depending on the personality one is attempting to model.

Unlike much of the work on modeling personality that has focused on surface behavior, a model of explanatory style attempts to characterize differences in underlying perceptions and thoughts that motivate behavior. Hayes-Roth, et al. developed synthetic actors showing relevant behavioral tendencies with respect to their personalities [2]. Gebhard, et al. adjusted the intensity of an agent's emotion based on personality traits [3]. Pelachaud, et al. introduced Greta, as a conversational agent, assigning different degrees of importance of certain goals according to its personality [4]. Paiva, et al. regulated an agent's emotional threshold and decay rate in accordance with the personality in an interactive system, called FearNot! [5]. While they have modeled an agent's different behavioral tendencies according to the agent's personality, they have hardly considered an agent's dispositional differences in understanding of social events based on the personality. Thus, we aim to concretize how an agent appraises social situations differently according to personality factors and how the appraisal differences influence the agent's emotional abilities to cope with the situations.

In this paper, we begin by introducing psychological explanatory styles and individual differences on the styles according to personality. Then, we present how to recast these theoretical explanatory styles to a computational framework. We also develop an interactive virtual environment that uses the model to allow participants to explore individual differences in the explanation of social events, a step toward the ultimate goal of developing applications that encourage the advancement of perspective taking and emotion-regulatory skills. Finally, we summarize our work and discuss future researches.

2 Explanatory Styles for Appraisals

Research about individual differences in expressional, logical, and emotional aspects has been studied extensively in psychology. Most psychological approaches structure personality in terms of abstract traits such as extroversion or neuroticism – e.g., The Big-Five model [6]. Traits are abstract constructs that have broad impact over many aspects of cognition and behavior, and an individual is characterized as some combinations of different levels of intensity of different traits. On the other hand, some psychologists have studied specific personality differences in greater detail, attempting to elucidate the underlying factors that produce these differences – e.g., explanatory styles [1].

In this paper, our goal is to make an agent understand situations differently according to its own personality. To achieve the goal, we base our studies on psychological explanatory styles, especially the work of Peterson and Seligman [1]. They define an explanatory style as a cognitive personality variable that reflects how people habitually explain the causes of events. They insist that a person's mental and physical health is affected by the person's explanatory style in important ways. Explanatory styles are closely associated with clinical disorder, e.g. depression, and help to predict whether a person will succeed in a wide variety of tasks [7] [8]. In addition, explanatory styles are straightforwardly investigated through several measurements, e.g., Attributional Style Questionnaire (ASQ), Content Analysis of Verbatim Explanation technique (CAVE), Expanded ASQ, etc [9]. The measured styles are exploited for cognitive behavioral therapy helping to improve the confidence and well-being of individuals.

To explain events, people basically answer the following questions: Who causes the situation? How long will the situation last? How much of my life does the situation affect? That is, explanatory styles are differentiated by three factors: Personalization, permanence, and pervasiveness[†]. *Personalization* shows the extent to which the explanation is internal ("it's me.") versus external ("it's someone else."). *Permanence* indicates a stable event ("it will last forever.") versus an unstable event ("it's short-lived"). *Pervasiveness* denotes an event as global ("it's going to affect everything that happens to me") versus specific ("it's only going to influence this").

People differ in their habitual explanatory tendencies based on their own personalities [1]. Especially, the tendencies clearly can be differentiated with pessimists or optimists. Pessimists have negative explanatory styles to explain events in their lives. They believe that negative events are caused by them (internal), always happens (stable), and affect other all areas in their life (global). They see that positive events are caused by things outside their control (external), probably will not happen again (unstable), and are isolated (specific). In contrast, optimists have positive explanatory styles. They explain negative events as not being their fault (external), and consider them as being isolated (unstable) that have nothing to do with other areas of their lives or future events (specific). They consider positive events as having happened because of them (internal). They see them as evidence that more positive things will happen in the future (stable), and in other areas of their lives (global).

[†] We quote terms indicating key factors of an explanatory style from Seligman's book [10].

3 Computational framework for explanatory style

In recasting psychological explanatory styles to a computational framework for socially interactive agents, we model core conceptual variables underlying the explanatory styles. To make the agents present different explanatory tendencies on situations according to their personalities, we design the algorithm to regulate degrees of explanatory variables based on the personality factors. Then, we specify how the variables have influence on the agent's explanatory process and the assignment of emotional states. To provide a solid framework for modeling differences in explanatory styles, we base our work on a computational appraisal theory of Gratch and Marsella [11], and especially theoretical developments on modeling social attribution [12]. Thus, an agent appraises the significance of events in its environment in terms of its relationships to its beliefs, desires and intentions. Then, the agent explains the situation based on the assessment, and reflects the explanation to emotion selection about the circumstance. Fig. 1 shows an overview of our computational framework.

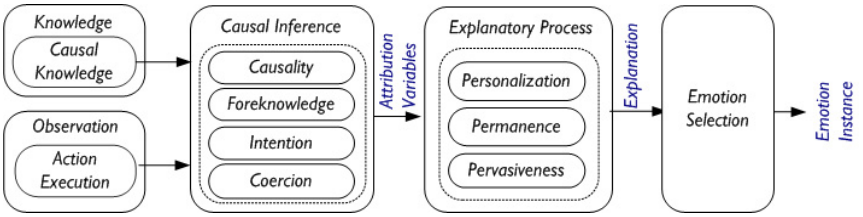


Fig. 1. An overview of our computational framework. An agent infers causal information about an event, explains the event according to its explanatory style, and changes an emotional state by reflecting the explanation.

3.1 Causal Inference

To make an agent infer causal information about an event, we need to represent the agent's mental state concerning actions and states [11]. An action consists of a set of preconditions and effects, and is associated with a performer (an agent that performs the action) and an authorizer (an agent that possesses the authority over the action). For example, if a student wishes to use the toilet, and needs to ask the teacher for permission, the student is the performer and the teacher is the authorizer. In addition, each state can be assigned a numerical value in the interval [-100, 100] denoting the agent's preference (utility) for the state. In our approach, the preference implies how much the state contributes to achieve the goal [13]. Thus, a state associated with positive value of preference is desirable for helping the agent accomplish its intended goal. In addition, the relationship between actions and states is represented by causal establishment or threat relation, i.e. the effect of action can establish or threaten the goals. A plan to achieve the intended goal is composed of a set of actions, states and their relationships. Moreover, each state is appraised, in accordance with

computational appraisal theory in terms of appraisal variables: Relevance, likelihood, controllability, changeability, etc., and can result in an emotional response (see [11]).

An agent deduces causal information, i.e., causality, foreknowledge, intention, and coercion, about circumstances from causal evidence in social interaction. The agent judges who causally contributes to the occurrence of an event, and whether the agency has foreknowledge about the event. It also decides if the outcome is coerced or intended by some other agents. In this paper, it is beyond the scope of this paper to describe algorithms on how to infer the causal information. However, detailed axioms and inference algorithms can be found at [12].

3.2 Explanatory variables

Psychological explanatory styles have considered human's habitual dispositions on explanations about their situations through three key variables: Personalization, permanence and pervasiveness. In this paper, our goal is to build a computational model describing explanatory styles of socially interactive agents. Thus, we embody theoretical variables of explanatory styles into our model, and associate them with specific aspects of an agent's tendencies to appraise the situations.

Personalization refers to who causes a situation. It is closely related to the assignment of responsibility for the occurrence of an event. Especially, it is associated with the blame or the praise for the outcome. If an agent has internal personalization, it tends to blame or credit itself for the situation. Contrastively, if the agent externalizes the situation, it shows a tendency to explain the consequence by attributing the blame or the credit to some other agents or other external factors.

Permanence determines how long this situation will last. It has an effect on the appraisal of the persistence of an event. It is correlated with the assessment of controllability and changeability about a situation. Controllability is a measure of an agent's ability to control the circumstance. Changeability indicates how likely the situation will be changed without any intervention. Thus, if an agent thinks an outcome is persistent, the agent considers that the consequence is not changeable (low changeable) and the agent itself does not have any controllability (low controllable). On the other hand, the agent considers a variable circumstance as high changeable and high controllable.

Pervasiveness is a measure of how much a situation affects other aspects. It takes effect on judgments of other events. In our approach, it corresponds to an agent's appraisal biases for other circumstances. When an agent regards a previous effect as pervasive, it makes the agent hold a biased view. Accordingly, the agent evaluates other events toward similar appraisals of prior outcomes. For example, if an agent thinks of a bad circumstance as global, the agent tends to evaluate other consequences on negative lines. Meanwhile, if the agent considers the situation as specific, it does not show any influence on other appraisals.

3.3 Explanatory process

We have developed an explanatory process, as shown in Fig.2, which allows an agent to appraise a situation differently with respect to its own personality. We have extended Mao's framework for social explanations [12] to incorporate biases on a function of the agent's personality. We design to assign different tendencies on an agent's explanatory variables in accordance with personality factors. Thus, based on a dispositional personalization, an agent attributes responsibility for the occurrence of an event, and blames or credits for the circumstance in different ways. As the agent has different inclinations to evaluate the persistence of the situation, it assigns different degree of controllability and changeability of appraisals on the outcome. The agent also adjusts the extent of influence of previous circumstances by the degree of pervasiveness. Therefore, same situation can be evaluated differently according to the agent's different explanatory propensities.

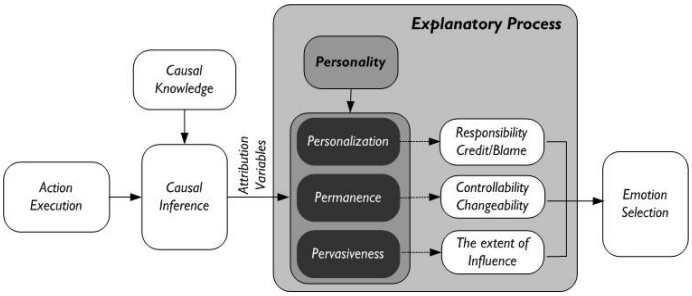


Fig. 2. An agent's different explanation depending on its explanatory tendencies. An agent differentiates to assign the responsibility, evaluate controllability and changeability, and determine the extent of influence of the situation based on the agent's personality.

According to an agent's explanatory characteristic on personalization, it differentiates the assignment of responsibility for the state. Furthermore, based on an agent's desirability on the circumstance, the agent blames or praises responsible agents for the situation. The assignment begins with a primitive action causing a set of effects. For an undesirable outcome, a pessimistic agent is biased to blame itself for the outcome. At first, the pessimistic agent judges whether it causally contributes to achieve the outcome or compels other agents to perform an action achieving the consequence. If the agent has any causality or coercion for the effect, the agent accuses itself for the undesirable state. On the other hand, an optimistic agent turns the responsibility of the negative outcome over to some other agents. So, when others who have causalities or coercions for the undesirable circumstance exist, the agent blames the consequence on them. In contrast, for a desirable effect, it shows an opposite way to assign the responsibility. That is, a pessimistic agent has a tendency to praise some other agents, such as indirect agencies or coercers, for the consequence. But an optimistic agent is apt to praise itself for the desirable outcome when the agent has causality or coercion on achieving the effect. For example, the project your friend is in charge of is a great success. If your friend has a pessimistic explanatory style, the friend applauds other teammates as they devoted time and

energy to the project. Contrastively, if your friend has an optimistic explanatory style, the friend takes credit to himself or herself in the success – e.g. self-admiration for good management of the project.

An agent's habitual tendency to assess the permanence of an effect is closely related to evaluate the persistence of the effect. Thus, the explanation is associated with appraisals of controllability and changeability for the circumstance. A pessimistic agent assigns high permanence (low controllability and low changeability) for an undesirable outcome, and low permanence (high controllability and high changeability) for a desirable outcome. Meanwhile, an optimistic agent attributes low controllability and low changeability for an undesirable state, and high controllability and high changeability for a desirable circumstance.

The agent's disposition on pervasiveness influences appraisal biases for other outcomes. In our model, it corresponds to the adjustment of the intensity of emotion instances associated with other appraisals. For an undesirable outcome, a pessimistic agent considers that the negative consequence affects all other appraisals. Thus, it increases the intensity of negative emotions (e.g. distress, shame, reproach, etc), while it decreases the intensity of positive emotions (e.g. joy, pride, admiration, etc) in other appraisals. Reversely, since an optimistic agent thinks of the undesirable consequence as isolated, it does not show any influence on other appraisals. However, when an agent has a desirable outcome, a pessimistic agent does not necessarily carry over the circumstance. Contrastively, as an optimistic agent regards the positive effect as pervasive, it increases the intensity of positive emotions and decreases the intensity of negative emotions in other appraisals.

3.4 Emotion selection

In our approach, another concern is how an agent's explanation influences on the assignment of an emotional state. The explanation contains information related to appraisal variables [11], especially desirability, controllability, and changeability, associated with an effect, and responsibility of the effect, the blame or the credit of the responsibility. Thus, we map the information into emotion instances based on OCC Model [14]. In OCC Model, responsibility has relevance to attribution emotions, e.g., pride, admiration, shame, reproach, etc. Accordingly, we define rules to assign the attribution emotions based on responsibility and the blame or the praise of the responsibility for an outcome. In addition, since desirability is related to assign the event-based emotions, e.g., joy, distress, etc, we list conditions for attributing the emotions. We append changeability and controllability to conditions for assigning the event-based emotions. Table 1 describes our basic principles to assign an emotion instance according to agent (*p*)'s perspective for the outcome *e*. Pride arises when *p* is responsible for producing a desired outcome *e.* Meanwhile, shame arises when *p* has the responsibility for causing an undesired state *e*. Respect arises when some other agent has the responsibility on achieving a desired state *e,* and *p* is praiseworthy for *e*. On the other hand, Reproach arises when some other agents are responsible for an undesired state *e*, and *p* is blameworthy for *e*. Distress occurs when agent *p* has low controllability in undesirable state *e*, which is seldom changed. Joy arises when *p* has a desirable state *e* which is unchangeable.

Table 1. Mapping explanations into emotion instances

Explanation configuration	Emotion instance
responsible agent(e) = p, causal attribution (p, e) = praiseworthy	Pride
responsible agent(e) = p, causal attribution (p, e) = blameworthy	Shame
responsible agent(e) = q ($\neq p$), causal attribution (p, e) = praiseworthy	Respect
responsible agent(e) = q ($\neq p$), causal attribution (p, e) = blameworthy	Reproach
desirability (p, e) < 0, controllability(p, e) = low, changeability (p, e) = low	Distress
desirability (p, e) > 0, changeability (p, e) = low	Joy

4 Implementation

We have developed an interactive environment that uses our model to allow participants to explore individual differences in the explanation of social events, a step toward the ultimate goal of developing applications that encourage the advancement of perspective taking and emotion-regulatory skills. As illustrated in Fig. 3, it makes participants experience flower gardening with a bluebird as a team. In this environment, there are two actors, gardener (participant) and guidance (bluebird), who worked as a team. A participant has an authority over a bluebird and orders commands, such as sprinkling water, etc, via a simple GUI. The bluebird actually carries out the commands in a virtual gardening environment, and then the virtual flower presents the effects of executed commands. Thus, the participants can learn the influence of the commands for flower gardening. Furthermore, the bluebird provides the participant with guidance through its own emotional responses to the status of the virtual flower.

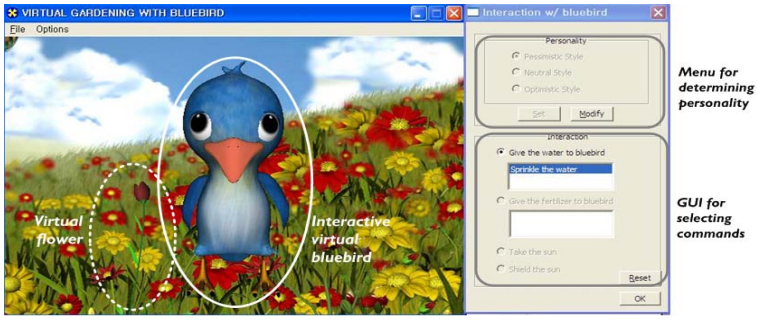


Fig. 3. An interactive environment with a bluebird. Participants can select a specific command through left GUI window. The bluebird executes the command and the virtual flower shows the effect of the selected command in this interactive environment.

We used our model to enable the bluebird to explain social situations in different ways depending on its personality factors. Before interacting with the bluebird, a

participant can predetermine the bluebird's explanatory style - e.g. pessimistic, neutral, and optimistic style. Then, selected style has an impact on the bluebird's explanatory tendencies for social interaction with the participant. In this environment, a participant has a *help-blooming* mission to achieve the goal *blooming-is-helped*. As shown in Fig. 4, there are two methods to achieve this: *apply-water* and *apply-fertilizer*. *Apply-water* consists of primitive actions; *give-water* and *sprinkle-water*, and *apply-fertilizer* is composed of *give-fertilizer* and *sprinkle-fertilizer*. *Sprinkle-water* and *sprinkle-fertilizer* have the effect *blooming-is-helped* which is a desirable goal to a bluebird and a participant. However, *sprinkle-fertilizer* has an undesirable side effect for the bluebird, which is that *root-becomes-weak*.

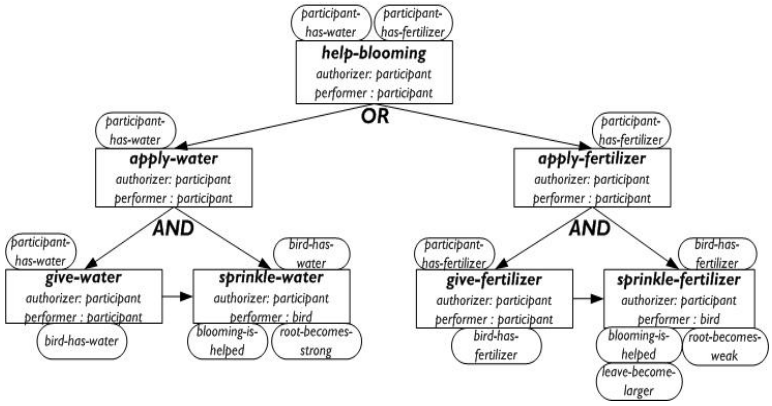


Fig. 4. A task structure of *help-blooming* in our interactive environment.

Let's imagine that a participant predetermined a bluebird's explanatory style as optimistic and coerced the bluebird to perform *sprinkle-fertilizer*. Then, an undesirable outcome *root-becomes-weak* occurred. Fig. 4 shows how our computational model informs the bluebird's explanation for the undesirable effect. Firstly, a bluebird knows that it is a causal agency and a participant is an indirect agency for the effect. The bluebird infers that the participant has foreknowledge about the effect and intends to achieve the consequence because the participant coerced the bluebird to perform *sprinkle-fertilizer* causing the negative outcome. According to a bluebird's optimistic explanatory style, the bluebird externalizes an undesirable state. Thus, it finds some other blameworthy agents, e.g., indirect agency, coercer, etc. Because the outcome is forced by a participant, the bluebird attributes the responsibility to the participant. Moreover, since the bluebird regards *root-becomes-weak* as unstable, it attributes high controllability and high changeability to the state. As the bluebird thinks of the negative outcome as isolated, it does not have any influence on other appraisals. Finally, the bluebird reproaches the participant for being blameworthy on the undesirable state *root-becomes-weak* according to our principles for determining an emotional state.

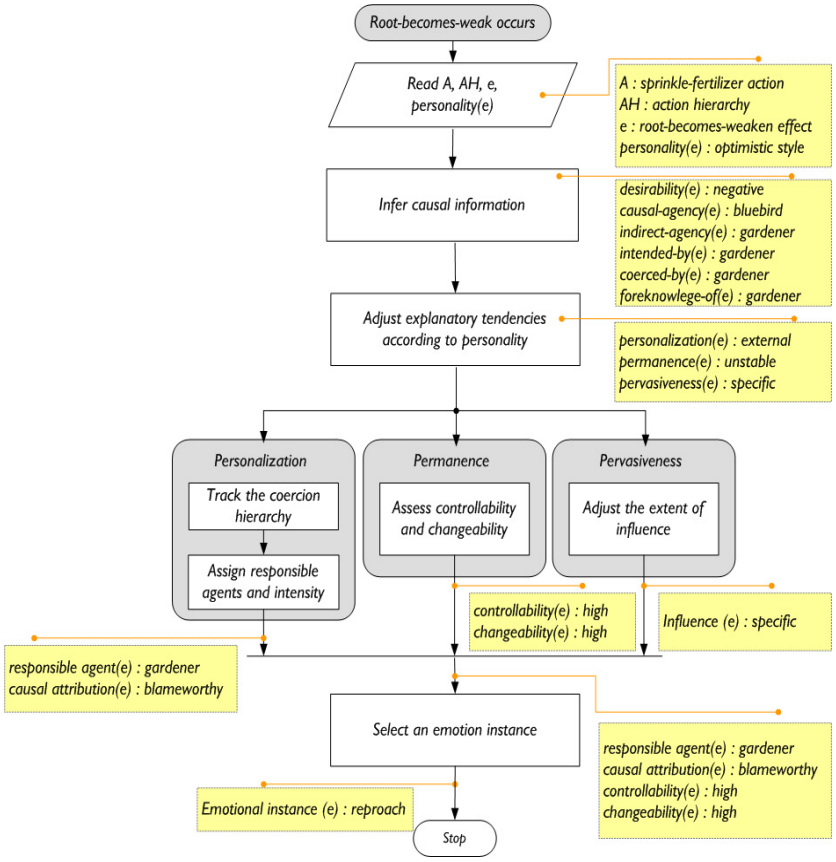
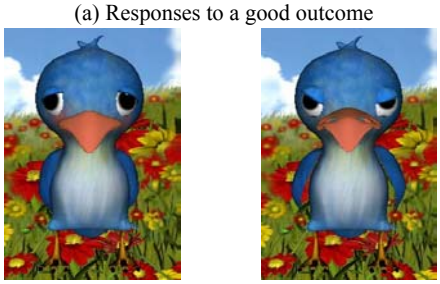


Fig. 4. An optimistic bluebird's explanation about a negative outcome *root-becomes-weak*.

Fig. 5 shows examples of a bluebird's different emotional responses to same events according to its explanatory tendencies. Since participants can interact with a bluebird as a team, it enables them to explore other team members' different explanatory styles when there is teamwork in the interactive environment. As a result, we can study how different explanatory styles of teammates influence the other teammate's performance of teamwork in team task environments. Ultimately, we can expect possibilities of applications that encourage the advancement of participants' perspective taking and emotion-regulatory skills in interactive environments.

(a) Responses to a good outcome



(b) Responses to a bad outcome

Fig. 5. Different emotional responses to same situation depending on its explanatory style. Left figures show a pessimistic agent's expression to a good or bad outcome, and right ones describe an optimistic agent's expression.

5 Summary and future work

In this paper, we presented a computational framework which allowed an agent to exhibit different explanatory tendencies for social events according to personality. Building on the framework, we modeled key conceptual variables underlying psychological explanatory styles, and designed to assign different explanatory tendencies depending on an agent's personality. We also specified how the variables inform the agent's explanatory process and the assignment of emotional states in social situations. Finally, we built an interactive virtual environment that used our framework to allow participants to explore individual differences in the explanation of social events, with the goal of encouraging the development of perspective taking and emotion-regulatory skills.

This work is still in its early stages. The current framework has focused on simple common sense rules which are sufficient and efficient for our practical application. Future research must extend our framework with more general rules for pervasiveness in explanatory styles. Since the implemented bluebird limits to exhibit its explanation through facial expression and simple movement, it is not enough to allow participants to understand a bluebird's explanations about social events. Therefore, we are planning to combine additional modalities, e.g., sound, etc, to improve participants' understanding about the bluebird's explanation. We are also planning to evaluate the

effectiveness of our proposed framework through a comparative study with other research. Furthermore, we will measure how participants' explanatory tendencies influence their comprehensions of the bluebird's explanations of social events in interactive edutainment systems.

References

1. Buchanan, G. Seligman, M.E.P.: Explanatory Style. Hillsdale, N.J.: Erlbaum, (1995)
2. Rousseau, D., Hayes-Roth, B.: A Social-Psychological Model for Synthetic Actors. Agents' 98 (1998) 165 – 172
3. Gebhard, P.: ALMA – A Layered Model of Affect. AAMAS'05 (2005) 29-36
4. F de Rosis, C Pelachaud, I Poggi, V Carofiglio and B De Carolis: From Greta's mind to her face: modeling the dynamics of affective states in a conversational embodied agent. International Journal of Human-Computer Studies (2003) 81-118
5. Paiva A., Dias J., Sobral D., Aylett, R., Zoll C., Woods, S.: Caring for Agents and Agents that Care: Building Empathic relations with Synthetic Agents. AAMAS'04 (2004) 194-201
6. Digman, J.: Personality Structure: Emergence of the Five-Factor Model. Annual Review of Psychology, Vol. 41 (1990), 417-440
7. Alloy, L.B., Peterson, C., Abramson, L.Y., Seligman, M.E.P.: Attributional style and the generality of learned helplessness. Journal of Personality and Social Psychology, Vol. 46 (1984) 681-687
8. Peterson C., Vaidya R.S.: Explanatory style, expectations, and depressive symptoms. Personality and Individual Differences, Vol. 31 (2001) 1217-1223
9. Fernandez-Ballesteros, R.: Encyclopedia of Psychological Assessment. Sage (2002)
10. Seligman, M.E.P.: Learned optimism: How to change your mind and your life. New York: Random House (1998)
11. Gratch, J., Marsella, S.: A Domain-independent framework for modeling emotion. Journal of Cognitive Systems Research, Vol. 5, Issue 4 (2004) 269-306
12. Mao, W.: Modeling Social Causality and Social Judgment in Multi-agent interactions. Ph.D Dissertation (2006)
13. Gratch, J., Marsella, S.: Technical details of a domain independent framework for modeling emotion. from www.ict.usc.edu/~gratch/EMA_Details.pdf
14. Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotion. Cambridge, UK Cambridge University Press (1988)