Affective Expressions in Conversational Agents for Learning Environments

Effects of curiosity, humour, and expressive auditory gestures

by

Jessy Ceha

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Computer Science

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Ehud Sharlin
Professor, Department of Computer Science
University of Calgary

Supervisor: Edith Law
Associate Professor, Cheriton School of Computer Science
University of Waterloo

Internal Members: Edward Lank
Professor, Cheriton School of Computer Science
University of Waterloo

Kerstin Dautenhahn
Canada 150 Research Chair in Intelligent Robotics
Professor, Departments of Electrical and Computer Engineering/
Systems Design Engineering,
Cross-appointed with the Cheriton School of Computer Science
University of Waterloo

Internal-External Member: Elizabeth Nilsen
Interim Assistant Vice President,
Graduate Studies and Postdoctoral Affairs
Professor, Department of Psychology
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This dissertation includes first-authored peer-reviewed material that has appeared in conference proceedings published by the Association for Computing Machinery (ACM). The ACM's policy on reuse of published materials in a dissertation is as follows[1]:

> *"Authors can include partial or complete papers of their own (and no fee is expected) in a dissertation as long as citations and DOI pointers to the Versions of Record in the ACM Digital Library are included."*

The following list serves as a declaration of the Versions of Record for works included in this dissertation.

Portions of Chapter 3
Jessy Ceha, Nalin Chhibber, Joslin Goh, Corina McDonald, Pierre-Yves Oudeyer, Dana Kulić, and Edith Law (2019). *Expression of Curiosity in Social Robots: Design, Perception, and Effects on Behaviour.* In Proceedings of CHI Conference on Human Factors in Computing Systems (CHI '19), May 4–9, 2019, Glasgow, Scotland UK. ACM, New York, NY, USA, 11 Pages.
DOI: 10.1145/3290605.3300636

Portions of Chapter 4
Jessy Ceha, Ken Jen Lee, Elizabeth Nilsen, Joslin Goh, and Edith Law (2021). *Can a Humorous Conversational Agent Enhance Learning Experience and Outcomes?* In Proceedings of CHI Conference on Human Factors in Computing Systems (CHI '21). ACM, New York, NY, USA, Article 685, 1–14.
DOI: 10.1145/3411764.3445068

[1] See https://authors.acm.org/author-services/author-rights. Accessed September, 2021.

Abstract

Conversational agents – systems that imitate natural language discourse – are becoming an increasingly prevalent human-computer interface, being employed in various domains including healthcare, customer service, and education. In education, conversational agents, also known as pedagogical agents, can be used to encourage interaction; which is considered crucial for the learning process. Though pedagogical agents have been designed for learners of diverse age groups and subject matter, they retain the overarching goal of eliciting learning outcomes, which can be broken down into cognitive, skill-based, and affective outcomes. Motivation is a particularly important affective outcome, as it can influence what, when, and how we learn. Understanding, supporting, and designing for motivation is therefore of great importance for the advancement of learning technologies.

This thesis investigates how pedagogical agents can promote motivation in learners. Prior research has explored various features of the design of pedagogical agents and what effects they have on learning outcomes, and suggests that agents using social cues can adapt the learning environment to enhance both affective and cognitive outcomes. One social cue that is suggested to be of importance for enhancing learner motivation is the expression or simulation of affect in the agent. Informed by research and theory across multiple domains, three affective expressions are investigated: curiosity, humour, and expressive auditory gestures – each aimed at enhancing motivation by adapting the learning environment in different ways, i.e., eliciting contagion effects, creating a positive learning experience, and strengthening the learner-agent relationship, respectively.

Three studies are presented in which each expression was implemented in a separate type of agent: physically-embodied, text-based, and voice-based; with all agents taking on the role of a companion or less knowledgeable peer to the learner. The overall focus is on how each expression can be displayed, what the effects are on perception of the agent, and how it influences behaviour and learning outcomes. The studies result in theoretical contributions that add to our understanding of conversational agent design for learning environments. The findings provide support for: the simulation of curiosity, the use of certain humour styles, and the addition of expressive auditory gestures, in enhancing motivation in learners interacting with conversational agents; as well as indicating a need for further exploration of these strategies in future work.

Acknowledgements

There are a number of people who made this thesis possible. To begin, I would like to thank my supervisor, Edith Law, for her guidance, for the countless opportunities she provided, and for her continued support of my various research and career endeavours. Thank you to Dana Kulić for introducing me to the field of human-robot interaction, as it inspired me and helped me find my footing when I started this process. A special thanks to the mentors I met along the way as well, including Pierre-Yves Oudeyer, Joslin Goh, and Elizabeth Nilsen. The time, respect, and advice they offered were significant to my PhD journey.

I am grateful to my Examining Committee for sharing knowledge from their respective fields, and suggesting valuable critiques to improve my thesis. I also wish to acknowledge all of the participants that took part in my studies, whose time and energy were vital in making each research project a success.

I would like to express my gratitude to my co-authors and collaborators, who were a pleasure to work with, and from whom I learned so much. To the lab-mates and colleagues that I started this adventure with, and to those I met throughout – thank you for making my time at Waterloo an enjoyable and unforgettable experience.

Lastly, I could not have gotten this far without the love and support of my family. I will be forever grateful for their unwavering belief in me.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

Conversational agents, or systems that imitate natural language discourse, have become a prevalent human-computer interface. Conversational agents can provide task-based, social, and companion-based interactions through text, speech, and multi-modal communication. Text-based agents, considered the most basic form, commonly communicate with users through a simple text interface, while voice-based agents communicate largely through just a voice. Embodied agents, on the other hand, have a simulated virtual or physically-present character resembling a human or other form, and are able to additionally contribute gestural and emotional expressions, through such things as body language, facial expressions, and vocal characteristics.

In education, students today are encouraged to actively construct their own knowledge and understanding, and evolve their knowledge base through interactions with others [131]. This model of education is based on the theories of *social constructivism*, according to which human development is social in nature and an important factor in the learning process [160], and conversational agents provide a unique opportunity to encourage this type interaction within computer-mediated learning environments. The overarching goal of pedagogical agents are to elicit learning outcomes. Learning outcomes can be considered as changes in the following areas: cognitive, skill-based, and affective [80]. It is common in pedagogical agent research for effects to be present in one, two, or all three of the domains of learning outcomes, especially if taking into account affective outcomes, as affect has been found to play an important role in human learning and decision-making [13, 55]. Motivation – to be moved to do something – is an especially important affective outcome when it comes to learning [138], and as such, researchers are exploring the factors that can optimize learner motivation.

Pedagogical agent researchers have explored various features of the design of agents, including gender, competency, interaction type, ethnicity, multiplicity, and feedback, to explore what effects they have on different learning outcomes. A crucial aspect of human conversation are social cues, and research on conversational agents indicates that endowing agents with the ability to express various social behaviours can lead to increases in such things as usability, efficiency, trust, and rapport [105]. In terms of pedagogical agents, growing evidence suggests that social behaviours (e.g., exhibiting social intelligence [163]) can also enhance various learning outcomes, however there still exist open questions as to *which* behaviours should be implemented and *how* they should be implemented — with many possible designs to be explored.

One specific type of social cue is the ability of an agent to express or simulate affect. As mentioned previously, affect plays a crucial role in learning, but affective expressions can also influence the interaction between human and agent, or learner and agent in an educational context, possibly because it gives insight into the internal state of the agent. Through affective expressions, research suggests that agents can modify the learning environment in such a way that is hypothesized to be beneficial for the learner, thereby enhancing learners' own affective states – such as engagement, motivation, and feelings of responsibility – as well as cognitive learning outcomes.

This thesis explores three affective expressions, in particular: **curiosity**, **humour**, and **expressive auditory gestures**. Each expression was chosen informed by prior work across multiple domains, as they are hypothesized to enhance learner motivation and/or cognitive outcomes by adapting the learning environment in the following ways: eliciting contagion effects of the simulated affective state, creating a positive learning experience, and strengthening the learner-agent relationship, respectively.

Eliciting contagion effects – "the tendency to automatically mimic and synchronize expressions, vocalizations, postures, and movements with those of another person's and, consequently, to converge emotionally" [66]. The impacts of contagion in interaction with conversational agents is still being explored, but studies indicate promising results for educational contexts (e.g., [121]). Therefore, using the idea of eliciting contagion effects, how can learner motivation be enhanced? **Curiosity** – the desire to explore or obtain unknown information [96] – is a form of intrinsic motivation (doing an activity for its inherent satisfaction rather than for some external goal), and has been shown to have numerous benefits for learners. Much research on curiosity has focused on its elicitation through stimuli that are novel, surprising, conceptually conflicting, or uncertain [144, 22, 86]. However, recent research has begun to explore whether pedagogical agents can foster curiosity in learners by expressing it themselves. For example, to mimic an intrinsically motivated (curious) learner, agents have been designed to add both verbal and nonverbal curiosity-driven behaviours to their dialogue and interaction, i.e., justification, argumentation, and hypothesis verbalization [120], free exploration and uncertainty seeking [56], and selecting 'interesting' objectives based on prior experiences and knowledge [165, 167]. In general, curious agents have been shown to enhance curiosity in the learner, but results do not always support an increase in cognitive learning gains, and the curious agents sometimes differ from the control along multiple dimensions, making it difficult to identify what influenced the observed changes. Moreover, not all curiosity-driven behaviours have been explored yet.

Creating a positive learning experience. Researchers propose that a positive learning experience can influence learning outcomes, including motivation, by providing an opportunity for cognitive rest, increasing engagement, and providing memory cues. **Humour** is an integral part of human communication and can be expressed verbally and non-verbally, intended to elicit laughter and mirth. In human-human teaching, material taught with humour can lead to better retention and recall [34, 29, 172, 50], increase learner interest and ability to engage in divergent thinking [41, 172, 171], and has been found to enhance self-esteem and motivation [17, 54]. Research investigating the effects of humour in conversational agents suggests that different forms of humour can enhance how a social robot is perceived [76], and improve perception of task enjoyment, robot personality, and robot speaking style [110]. In interactions with virtual agents, conversational and situation-specific jokes have been found to affect how cooperation is perceived in an agent [83], humour has been proposed as a means of recovering from error situations while providing a pleasant user experience [111], and humour has been shown to significantly motivate healthy behaviours [117]. Considering the various types of humour, in terms of expressing affect, humour styles – behavioural tendencies related to the uses or functions of humour in everyday life – have been found to be correlated with various moods. Although some work has explored the expression of humour in conversational agents, the use of humour by pedagogical agents, and in particular humour styles, is less explored.

Strengthening the learner-agent relationship. Learners that feel more responsible or have a better relationship with their agent, are suggested to be more motivated to put in effort while interacting with their agent [89, 115], and research outside of educational contexts suggests that affective expressions of various forms can enhance the agent-human relationship, e.g., modifications of voice pitch were found to improve interaction quality with a social robot [110]. Many studies have explored expressivity through such behaviours as facial cues [11], utterance-level prosody modifications [97], and bodily postures or gestures [121], but a lesser studied form of expression are **expressive auditory gestures**, even though similar to visual cues, they are able to convey information about affective state [14]. In human-human conversation, a form of vocal gesture, known as *interjections* (e.g., “ah”, “aww”, “blah”, “eww”), are used to communicate a speaker’s emotions and attitudes, and expressive interjections in particular, express reactionary feelings or emotions, as well as feelings that result from what one comes to know or understand. And although interjections are common in human dialogue, and are suggested to strengthen relational bonds between humans and agents [136], questions remain as to how users perceive an agent that adds interjections to communicate its affective state, and what effects these gestures have on the interaction in an educational context. Moreover, in contrast to humans, agents can express their emotions and attitudes through interjections but also other types of sounds. One expressive type of sound is music, and *musical expressions* have a number of similarities with the human voice in conveying internal states. Some work has investigated the addition of musical expressions and other sounds to conversational agent speech, looking at: how they should be designed, whether they can be perceived as intended, and what the effects are on perception of the agent (e.g., [69]), however not a lot of research has looked at adding interjections or musical expressions to pedagogical agents and the impacts on relationship building and learning outcomes.

1.2 Thesis Statement

Previous research indicates that pedagogical agents can be beneficial for learning, and that the design of agents plays a role in their success. It is therefore crucial to carefully consider and develop an understanding of the features that can play a role in affecting learning outcomes when implemented in learning environments. Notably, agents using social cues are suggested to adapt the learning environment in such a way that enhances affective and cognitive learning outcomes. One social cue that is suggested to be especially relevant for enhancing learner motivation is the expression or simulation of affect in the agent itself. The thesis is therefore structured around the following statement:

> *Pedagogical agents can use affective expressions to modify the learning environment, changing learners’ perceptions, enhancing affective outcomes such as motivation, and positively influence cognitive outcomes, such as recall of the material.*

In particular, three affective expressions are explored – curiosity, humour, and expressive auditory gestures. Each expression is hypothesized to change the nature of the learning environment in a different way so as to enhance motivation: by eliciting contagion effects, creating a positive experience, and strengthening the learner-agent relationship, respectively.

To explore this, the thesis addresses the following research questions:

> **RQ1:** What are the effects of having a pedagogical agent convey intrinsic motivation (i.e., curiosity) to learners using verbal expressions?
>
> **RQ2:** What are the effects of having a pedagogical agent express humour, using different humour styles?
>
> **RQ3:** What are the effects of adding expressive auditory gestures (i.e., interjections and musical executions) to the speech of a pedagogical agent?

In order to gain a thorough understanding of the impacts of the various affective expressions on learning outcomes, each study focuses on the following main effects:

> **Perception.** How do learners perceive/identify the implemented expression and how does the expression influence learners' framing of the agent — with respect to such measures as: likeability, perceived intelligence, rapport, etc.?
>
> **Emotion and Behaviour.** What are learners' emotional and behavioural responses (with regards to motivation, etc.) to the expression?
>
> **Learning.** Are there any gains in cognitive learning outcomes as a result of the added expression?

To answer these questions, the thesis presents three studies in which each expression is carefully designed and implemented in a separate agent, physically-embodied, text-based, and voice-based. Three agents were used as they are reflective of the current technologies on the market, and each offers different expressive capabilities. Further rationale behind the use of each agent is explained in the corresponding chapters.

The agent in each study took on the role of a companion or less knowledgeable peer to the learner. A learning companion agent that learns with, and is taught by, the learner, is based on the concept of "learning-by-teaching" (e.g., [24]), and are of particular interest in education, as learning-by-teaching can be a more enriching experience than learning by oneself [43]. In contrast to a teacher agent, teachable/learning companion agents are found to prompt learners to pay more attention, reflect on misconceptions, and elaborate on explanations [132]. However, in human-human tutoring, tutee errors do not lead to more learning for the tutee [161] and thus teachable agents are able to capitalize on the beneficial effects for learners as peer tutors, while avoiding the detrimental ones of being a peer tutee. These agents present a unique perspective for exploring motivation enhancement as evidence suggests that when learners taking on the role of the more-knowledgeable partner in the dyad, feel more responsible or have a better relationship with their agent, are more motivated to put in effort to teach their agent, and as a result also learn more [89, 115].

The subject matter in each study revolved around learning how to classify rocks. The topic of geology and earth science is taught at elementary-school level to university-level. This means that open educational resources covering the topic were available to be used for the studies, and it allows for the content to be adapted to various age groups as required, increasing the generalizability of the task.

1.3 Contributions

The findings contribute to our understanding of using affective social speech signals in pedagogical agents. The results from the studies indicate that pedagogical agents using affective expressions – in particular, curiosity, humour, and expressive auditory gestures – can be designed in such a way that modify learner perceptions of the agent and learning experience, enhance motivation, but have no effect on cognitive outcomes. The results challenge prior work that suggest associations between affective expressions and cognitive learning outcomes, but rather provides evidence that agents using affective expressions may lead to more positive affective states that allow a learner to focus and be motivated, without leading directly to learning.

Furthermore, separately, the studies contribute a number of interesting findings for each expression and research area surrounding it. In terms of curiosity, various ways have been explored to foster this form of intrinsic motivation, but the study contributes to the understanding of eliciting curiosity through contagion during social interaction; the study on humour highlights the benefits and pitfalls of implementing affiliative and self-defeating humour styles in pedagogical agents; and the study on expressive auditory gestures provides a better understanding of how they can be used to influence rapport and learning outcomes.

The remainder of the thesis is organized as follows: Chapter 2 covers background literature on pedagogical agents and learner motivation, Chapters 3, 4, and 5 provide related work on each specific expression and describe the studies conducted, and Chapter 6 presents an overview of the most significant findings, and discusses insights, limitations, and directions for future work.

Chapter 2

Background Literature

2.1 Conversational Agents

Conversational agents – systems designed to communicate through natural language – have found applications across various fields, including museum tours (e.g., [23]), healthcare (e.g., [156]), and education (e.g., [57]). The terminology used to describe such systems differs across disciplines, including chatbots, virtual assistants, social robots, dialog systems, intelligent virtual agents, and embodied conversational agents, but in this thesis the term conversational agent will be used to cover any system designed to interact with humans by simulating natural language discourse. The natural language in conversational agents can be both voice or text, and the agents can take on various forms and ways of expression. Text-based agents, e.g., chatbots, commonly communicate through a text interface, with users typing or saying responses, or selecting from multiple choice answers. Voice-based agents, also known as spoken-dialog systems or voice-user interfaces, frequently implemented on personal computers, smartphones (e.g., Apple's Siri, Microsoft's Cortana), and smartspeakers (e.g., Google's Home, Amazon's Echo), communicate largely through voice alone. Embodied agents can be both virtual and physically present in the world, with the main distinction with voice-based agents being embodiment – defined as having a bodily presence or visual representation (as a human or other form ranging from cartoon-like to highly detailed), that can display various multimodal verbal and nonverbal behaviours.

2.2 Conversational Agents in Education

Learning technologies started largely with intelligent tutoring systems (ITSs). These systems were based around evidence showing that students learned significantly better from individual tutoring than classroom instruction [26]. ITSs are designed to provide timely feedback, support active engagement in learning, cater difficulty to the zone of proximal development, and help students through cognitive disequilibrium, by inferring the cognitive states of the learner. There are however certain instructional strategies that are based in dialog: asking-deep reasoning questions (e.g., why, how, what if), self-explanations, collaborative interactions, and fostering common ground and terminology. Verbal discourse specifically provides the opportunity to express and communicate knowledge in different ways to aid learning, such as through self-reflection, answering deep questions, generating questions, and resolving conflicts in understanding [114]. Conversational agents are able to provide this verbal discourse to learners.

Initial studies on pedagogical agents were heavily influenced by the work of Reeves and Nass [128] on human-computer interaction. They posited the *Media Equation* hypothesis, suggesting that humans respond to media (including computers) in the same ways they do other humans. Early researchers therefore investigated how the Media Equation might apply to learning technologies, resulting in works such as Herman the Bug [90], an agent that facilitated learning in an environment called Design-a-Plant. The researchers proposed that if humans attribute human-like characteristics to computers, then a computer expressing itself through an animated human-like agent may accentuate those effects—as an agent could emulate other aspects of human-human interaction including emotion and empathy. From their findings, the researchers revealed what they termed the *Persona Effect*—that an animated pedagogical agent with a life-like persona in an interactive learning environment can facilitate learning. Other early pedagogical agents include Cosmo [91], present in a virtual learning environment providing advice to students on Internet packet routing, and AutoTutor [114], a natural language tutoring system developed over the course of two decades. Cosmo used a behaviour planner based on the physical properties of the virtual world and used this information to create diectic gestures, motions, and utterances. The work provided a framework for endowing 'deictic believability' in animated agents. AutoTutor has been applied to the topics of computer literacy, conceptual physics, biology, and critical thinking. Early versions of AutoTutor were focused on examining the impact on deep learning of co-constructed explanations, feedback, conversational scaffolding, and subject matter content.

Beyond guiding students, pedagogical agents have also started to take on other roles such as co-learners or novices. *Co-learner agents* are based on approaches of cooperative learning, where collaboration develops a base for social communicative interactions and relationship development. *Novice/teachable agents* are based on the learning-by-teaching technique; a widely studied and practiced topic within the education domain. The approach is inspired by the *Protégé effect*, which demonstrates that learning for the sake of teaching others is more beneficial than learning for one's own self [25].

With advancements in various fields including natural language processing and machine learning, conversational agents are increasingly able to process and respond to complex user requests, as well as mimic human-human interaction through verbal and nonverbal behaviours to provide a natural interaction [48]. Along with the technical advancements, many researchers argue that conversational agents must act socially and display expressive behaviours to successfully interact with humans. Indeed, many studies show that human-like social cues have a significant impact on how users perceive and interact with agents, i.e., feelings of trust [21], satisfaction [88], and human-agent relationship [1]. These insights are similar to the *social agency theory*, a frequently applied theoretical framework for research on pedagogical agents [103] which posits that social cues can be used in learning environments to encourage learners to consider their relationship with a computer to be similar to interaction with a human, and thereby elicit deeper processing of the subject matter. A number of social cues have been implemented in pedagogical agent scenarios showing positive effects on learning outcomes, including off-task conversation [62], complementing, encouraging, and showing concern [98], goal alignment [4], social intelligence based on politeness theory [163], attentive backchanneling feedback [51], empathy and encouragement [87], social voice-adaptive speech [97] and argumentation [5], to name a few.

Another type of social cue is the expression of affect (feelings, moods, emotions). Prior work suggests that affective expressions are crucial for enhancing the communication between human and agent, which may in-turn positively influence learning outcomes. For example, Baylor and Kim [12] investigated whether three human instructional roles (Expert, Motivator, and Mentor) could effectively be simulated by pedagogical agents. To create each role, the researchers took into account agent image, animation, affect, and voice. The Expert agent was informative but displayed no affect, and had an authoritative and monotone voice. The Motivator and Mentor agents expressed emotions of acknowledgement, confusion, disapproval, excitement, pleasure, and surprise, but the Motivator agent was only encouraging, whereas the Mentor agent was both informative and encouraging. The Motivator agent's voice was effusive and enthusiastic, whereas the Mentor agent's voice was confident and calm. Through two controlled experiments, the researchers found that the agent roles were perceived as intended by students and that the Expert agent led to increased information acquisition, the Motivator agent resulted in increased self-efficacy, and the Mentor agent led to overall improved learning and motivation.

Researchers of pedagogical agents have explored a multitude of affective expressions. The expressions are often chosen so as to facilitate certain 'tactics' hypothesized to be beneficial for the learner; three of which particularly related to motivation are: eliciting contagion effects, creating a positive learning environment, and enhancing the human-agent relationship.

2.2.1 Eliciting contagion effects

The expression of affect is an important aspect in human conversation and can elicit contagion effects. Emotional contagion describes *"the tendency to automatically mimic and synchronize expressions, vocalizations, postures, and movements with those of another person's and, consequently, to converge emotionally"* ([66], p. 153-154). Emotional contagion has been shown to increase cooperation, decrease conflict, and improve perceived task performance in work group dynamics [9], as well as significantly influence amounts donated to charity [147]. Although much research on emotional contagion has been conducted in the social sciences domain, it's impact on interaction with conversational agents is still being explored, and studies indicate promising results for educational contexts.

Park et al. [121], programmed a fully autonomous peer robot (size and look of a teddy bear) with verbal and non-verbal behaviours suggesting it had either a growth mindset or a neutral mindset as it played puzzle solving games under time pressure, with a child (around 7 years old). The robot with a growth mindset chose more challenging puzzles and made growth mindset related comments about its own and the child's abilities and efforts throughout the session, e.g., "You are not afraid of a challenge. I like that!". Verbal statements were accompanied by non-verbal cues expressing excitement. The neutral mindset robot chose similar difficulty level puzzles as the child and made calm, factual statements, such as "You solved the puzzle", and nodded its head. The results showed that children who had played the game with the robot who had a growth mindset, self-reported a stronger growth mindset after interaction, and tried harder—indicated by more attempts made over time, demonstrating more resilience to failure—during a challenging puzzle task than the children who had played with the neutral mindset robot.

Another study investigated fostering creativity in children. Ali, Moroso, and Breazeal [6] implemented verbal and non-verbal behaviours in a collaborative social peer robot (Jibo) that were designed to express creativity. Fifty-one 6-10 year old children played a game with either the creative robot or a non-creative one. The results indicated that after interacting with the creative robot, children generated more ideas, explored more themes related to ideas, and came up with ideas that were more creative, compared to the children that had interacted with the non-creative one.

2.2.2 Creating a positive learning experience

Researchers propose that an agent's off-task social conversational abilities may influence learning outcomes due to the off-task conversation creating a positive learning experience. For example, Gulz, Haake, and Silvervarg [62] have shown that off-task social conversation (i.e., dialogue with no relation to the learning task, such as reassuring, cheering up, small-talk, and self-disclosure) with a pedagogical agent can lead to more enjoyable game play experience and higher learning gains. After learning to play one of several sub-games in a mathematics game aimed at training basic arithmetic skills, students were tasked with teaching it to a pedagogical agent. The agent was a text-based teachable agent, and the researchers compared an agent without off-task social conversation abilities, to one with. They found that middle- and high-achieving 12-14 year olds reported more positive game experience with off-task conversation, and high-achievers rated higher self-efficacy beliefs and had better learning outcomes. The authors proposed that these positive effects may be due to the off-task social conversation providing an opportunity for cognitive rest, increasing engagement, providing memory cues, and promoting trust and rapport-building between the learner and agent.

Off-task conversation can also involve expression of affect. Liew, Zin, and Sahari [93] designed a virtual tutor to express enthusiasm, conveyed through the tone of voice, constant smiling, a high level of animated movement and head-nodding during speaking, and enthusiastic remarks such as "I love programming" and "this is an interesting topic", and compared it to a neutral agent who's tone of voice was calm and pleasant, maintained a neutral facial expression, and had a low level of animated movement. The researchers found that the enthusiastic agent significantly enhanced university students' positive emotions, intrinsic motivation, affective perceptions of the agent and learning environment, and cognitive outcome.

2.2.3 Enhancing the learner-agent relationship

In human-human interaction, affective expressions not only improve understanding and communication, but they are also relevant for building and strengthening relationships. The relationship between learner and agent is particularly important when considering the perspective that learning can be considered a social, interpersonal collaboration [160]. In order to develop relationships, research on conversational agents similarly suggests the effectiveness of expressing affective states. For example, the emotional coloring of an utterance (i.e., the activation (active/passive), evaluation (positive/negative), and power (dominant/submissive) has been found to enhance feelings of rapport with a voice-based agent – i.e., the ability to relate to others in a way that creates a level of trust and understanding [1]; variations in voice pitch can result in increased ratings of interaction quality with a social robot [110]; and body gestures and speech content increase perceived personality of a virtual agent [109].

In educational contexts, there is evidence to suggest similar results. For example, Baylor [11] designed five facial expressions for: neutral, serious, happy, surprised, and sad states, and found the results to suggest that the facial expressions enhanced learners' perception of the agent and positively influenced cognitive learning outcomes.

2.3 Learning Outcomes

The overarching goal of pedagogical agents are to elicit learning outcomes. Learning outcomes can be considered as changes in the following three domains: cognitive, skill-based, and affective [80] (see Figure 2.1).

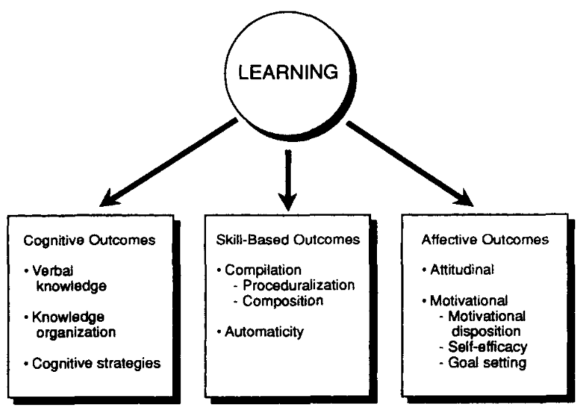


Figure 2.1: "A preliminary classification scheme of learning outcomes" from [80].

Cognitive outcomes are focused around the constructs of verbal knowledge (i.e., declarative knowledge), knowledge organization (i.e., mental models), and cognitive strategies (i.e., self-insight and metacognitive skills). Cognitive outcomes can be measured as amount and accessibility of knowledge, accuracy of recall, analysis of relation between different elements, and self-awareness and self-regulation. For example, Ramachandran, Huang, Gartland, and Scassellati [126] investigated how a robot tutor (NAO) may support metacognitive strategies during learning and complex problem-solving. They looked specifically at the think-aloud technique and how children could be instructed, reminded, and prompted by the agent during a problem solving task. They found that the robot and the think-aloud strategy promoted cognitive learning gains in the participants, and that the social robot tutor increased the students' engagement and compliance with the prompts to think-aloud.

Skill-based outcomes are concentrated on performance-type measures such as speed and fluidity of performance, error rates, generalization, but also attentional requirements and available cognitive resources. For example, Hood, Lemaignan, and Dillenbourg [67] conducted preliminary studies using the learning-by-teaching paradigm, by having students demonstrate handwriting to a teachable (NAO) robot, with the purpose of improving the students' own handwriting performance. Initial results show children engaging with the robot and improving its handwriting.

Affective outcomes include both attitudinal (i.e., direction and strength of attitude) and motivational (i.e., mastery vs. performance orientation (i.e., motivational disposition), perceived performance capability (i.e., self-efficacy), and goal setting, including level and complexity of goals, and goal commitment) outcomes. In other words, it includes the manner in which we deal with things emotionally, such as feelings, values, appreciation, enthusiasms, motivations, and attitudes [81]. For example, Lane et al. [84] showed that interactions with an animated pedagogical agent, 'Coach Mike', for informal computer science education, led to greater self-efficacy for computer programming when it expressed enthusiasm and used self-regulatory feedback.

It is common in research on pedagogical agents for effects to be present in one, two, or all three of the domains of learning outcomes, especially if taking into account affective outcomes, as affect plays an important role in human learning and decision-making, due to its influence on cognitive processes [13, 55].

Learner Motivation

Motivation – to be moved to do something [134] – is an especially important affective outcome as it can influence what, when, and how we learn [138]. Motivation is not a unitary phenomenon, people vary in their levels of motivation (i.e., how much) and in their orientation of motivation (i.e., the type of motivation). This orientation of motivation points to the underlying attitudes and goals that give rise to action. Deci and Ryan [39] put forth the Self-Determination Theory—distinguishing between different types of motivation, the most basic distinction being between: intrinsic motivation (doing something because of interest or enjoyment) and extrinsic motivation (doing something because of external prods, pressures, or rewards). In [134], Ryan and Deci argue for the importance of both types of motivation for successful learning. Although motivation has been defined a number of different ways over the years [139], researchers agree on certain behavioural indices that indicate the presence of motivation: task choice, effort, and persistence. Knowing the indicators, researchers are exploring the social and cognitive factors that can optimize learner motivation.

This thesis explores three affective expressions: **curiosity**, **humour**, and **expressive auditory gestures**, to adapt the learning environment through the three tactics mentioned previously, with the aim of enhancing motivation. The following chapters will discuss each affective expression separately in more detail, including how they are proposed to enhance learner motivation, and describe the studies conducted to explore their effects.

Chapter 3

Curiosity: Eliciting contagion effects

3.1 Related Work

Curiosity – the desire to explore or obtain unknown information [96] – is a form of *intrinsic motivation* and has been shown to enhance information-seeking behaviour [148] and improve memory retention (e.g., [60, 77]). Promoting curiosity also plays a role in adult [129] and infant [148] learning, and can foster early academic achievement, particularly for children with low socio-economic status [141]. It is therefore of interest to understand how to promote curiosity in learners. It is suggested that curiosity can be elicited through stimuli that are novel or conceptually conflicting, or feelings of uncertainty or surprise [19], and much research on curiosity, outside of educational contexts, have focused on its elicitation through stimuli that are novel, surprising, conceptually conflicting, or uncertain [144, 22, 86]. Another approach is to understand curiosity as it is shaped by the social environment, and recent research has begun to explore whether educational technologies like pedagogical agents can foster curiosity in learners, and consequently improve cognitive learning outcomes as well. One strategy that shows promise is to simulate curiosity in the pedagogical agent, and in so doing, elicit curiosity in the learner, by way of contagion.

Wu, Miao, and Shen [165] implemented a computational model of curiosity (based on Berlyne's extrinsic factors that stimulate curiosity: novelty, uncertainty, conflict, and complexity [20]) into a learning companion (embodied as a butterfly) based in a virtual environment, developed to teach students about plant transport systems. The model tracked the user's activity in the virtual environment to identify potentially interesting learning objectives, based on prior experiences and knowledge. The learning companion would then select these 'interesting' objectives (aka stimuli), thus mimicking a curious learner. In initial case studies with 10 primary-school students, the curious learning companion was found to help students explore more in the virtual environment. In [166], Wu, Miao, and Leung developed another computational model of curiosity, this time based on novelty, surprise, and uncertainty. Curiosity appraisal and learning were modelled in the agent to simulate curiosity. Curiosity appraisal involved the agent looking for knowledge gaps based on novel, surprising, or uncertain stimuli, whereas the learning process resolved the agent's curiosity by closing the knowledge gaps. Two learning companions were tested: a basic learning companion that was reactive and offered basic companionship during the game, and a curious learning companion with the computational model of curiosity, which offered basic companionship as well as learning and curiosity appraisal ability to direct the learners' attention to knowledge gaps. When the agent's curiosity was evoked, it would demonstrate curious behaviours by asking thought-provoking questions and encourage the participant to actively seek the answers. In comparison to the basic companion that did not learn with the participants, the curious agent was driven by curiosity to learn, and the agent automatically learned the student's mental model. The study with 31 university graduate students found that compared to the control condition, the curious learning companion resulted in more cognitive learning gains (measured by score difference between pre-test and post-test), was subjectively perceived as being better at facilitating learning, was more credible, human-like, and engaging, improved participants' self-efficacy (learners' belief about their competency), and made learners more interested (defined in the paper as: "a disposition organized through experience which impels an individual to seek out particular objects, activities, or goals for attention or acquisition [52]").

Another possible way of expressing curiosity is through verbal behaviour shown to evoke curiosity during human peer to peer dialogue. Sinha, Bai, and Cassell [145] presented the theoretical framework of curiosity, proposing ten verbal curiosity-driven behaviours exhibited during learning, which can positively influence curiosity in other group members – i.e., curiosity contagion. These include: (1) *Uncertainty:* Lack of sureness about something or someone, (2) *Argument:* Reasons, statements, or facts supporting or establishing a view point, (3) *Justification:* Providing information to make something clearer, (4) *Suggestion:* An idea or plan suggested by one group member, usually to get the other to do something, (5) *Agreement:* Group members' opinions or feelings are in harmony, (6) *Disagreement:* The opposite of agreement, (7) *Question Asking:* (on task vs. social) Any questions that are asked, not related to logistics of the game or task, (8) *Idea Verbalization:* Communication of an idea, (9) *Sharing Information:* Verbally communicating results, findings, or discoveries to the other player, and (10) *Hypothesis Generation:* Providing theories to explain something.

Gordon, Breazeal, and Engel [56] compared a curious virtual agent (on a tablet), a curious robot, and a non-curious robot in the context of supporting young children in learning how to read. The agents were portrayed as less-knowledgeable peers learning to read, prompting the children to teach new words. Two curiosity-driven behaviours were implemented in the curious conditions: free exploration and uncertainty seeking. The curious agents were enthusiastic about learning and exploration, challenged the child, and suggested novel moves in their co-player tablet app. Children's curiosity was measured by amount of information-seeking behaviour as a metric of free exploration (such as time taken to seek out a novel item), question generation (whereby participants were directed to ask as many questions about a topic as they could), and uncertainty seeking through "The Fish Task" [73], a tablet app recording users' choices of uncertain options. The curious robot resulted in significantly higher free exploration and uncertainty seeking than the non-curious robot, with no differences in question generation. Therefore, only those behaviours modelled by the robot had an impact on children's curiosity. Additionally, even though the children's curiosity was higher, the curious robot did not result in cognitive learning gains.

3.2 Study I

Guided by prior work, a study was designed to investigate how a pedagogical agent can convey curiosity to learners through on-topic question-asking, i.e., the agent was actively searching for new knowledge on the topic of the specific learning task, by asking questions, and whether those expressions in turn can affect the learners' own curiosity, and subsequently enhance cognitive outcomes. In contrast to prior work, only one curiosity-driven behaviour was implemented and was the only behaviour that was manipulated compared to a control, so that the curious agent did not differ from the neutral one along multiple dimensions, making it easier to identify what influenced any observed changes in curiosity. The curiosity-driven behaviour on-topic question-asking was chosen as it had not yet been explicitly implemented in a pedagogical agent as a means of expressing curiosity.

The focus was on understanding how agent curiosity is perceived, what impact it has on learning, and to provide a more detailed account of learners' curiosity – whether question-asking behaviour can lead to both behavioural and emotional curiosity contagion. Emotional curiosity contagion refers to the transfer of mood and affect (i.e., the emotional state of feeling curious), as opposed to behavioural curiosity contagion which is the transfer of behaviour indicative of curiosity (e.g., free exploration, uncertainty seeking).

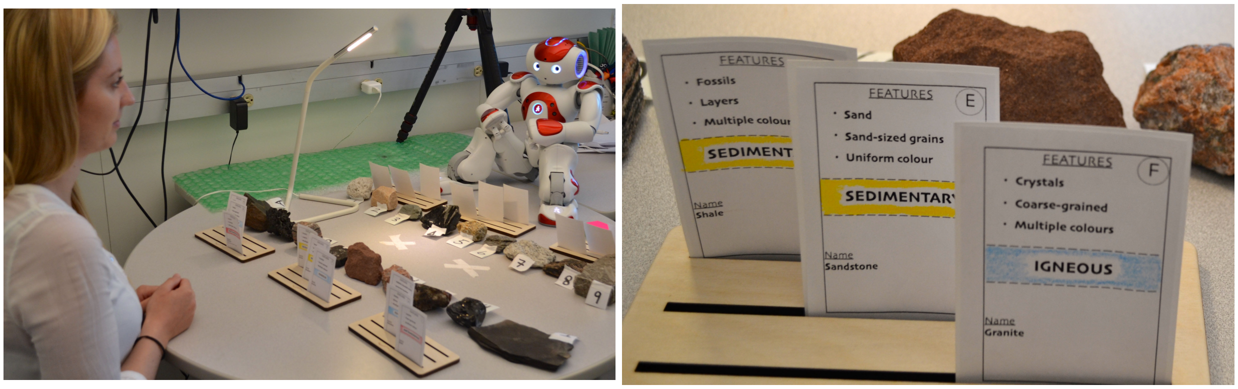
3.2.1 Agent

The agent chosen for this study was NAO, a small humanoid robot developed by SoftBank Robotics (shown in 3.1a). Robots have seen increasing deployment in educational settings [15]. In contrast to traditional technologies, students show more interest and higher performance while learning using educational robots [33, 8, 92], and compared to voice-only or virtual agents, the physical presence of robots can produce cognitive learning gains as well as more positive interactions [8, 92].

3.2.2 Task

As game play provides a unique opportunity to assess and support curiosity [73], LinkIt!, a novel educational game was designed for the interaction in this study. LinkIt! is played between a robot and student, and was designed to teach classification of rocks into rock types (sedimentary, metamorphic, or igneous) based on visual features, such as fossils. Players (robot and student) sit opposite each other, with a row of 9 rocks in front of each of them (3 sedimentary, 3 metamorphic, and 3 igneous, in random order; setup shown in Figure 3.1a). Every rock in a player's row has a card associated with it, containing the 'ground truth' (Figure 3.1b), i.e., three visual features, the type, and the name, which are not visible to the other player. The features were determined in consultation with geologists and are therefore standard descriptors of the rocks.

The aim of the game is for players to *link* each other's rocks by finding those with a common feature or type. Each round starts with the student taking the top card from an upside down deck placed in between the players. This card contains either a feature or rock type. During the round, both the robot and the student try to identify one of the other player's rocks which they think has this property. Once a player makes a guess, the other player silently reads the associated ground truth card to check whether the guess is correct. If both players are correct in identifying the property in each other's rocks, they have successfully linked them. Once linked, players are able to talk about the other features, rock types, or names that are listed on the ground truth cards for the linked rocks. The rocks can then be placed to the side of the game area, and a new card is pulled from the upside down deck. If a player guesses incorrectly, the other player can give them a hint, before trying again. If they guess incorrectly a second time, the round is over, and they will have to pick a new card from the upside down deck. The game ends once all cards in the upside down deck have been played, or all rocks have been linked.



(a) (b)

Figure 3.1: a) Setup of the game LinkIt!, designed to study human and robot behaviour, and played by participants in the study. b) The ground truth cards used in the game.

3.2.3 Experimental Conditions

There were three conditions in the study: 1) *Curious+Reveal* – the robot is curious and reveals the reason for its curiosity, 2) *Curious* – the robot is curious but does not reveal the reason, and 3) *Neutral* – the robot is not curious. As effective teaching requires an understanding of the learning progress of the student, and not being able to figure out the internal state of an agent can lead to a loss of trust, satisfaction, and acceptance [108], the Curious+Reveal condition was included to investigate whether providing explanations as to why or why not the agent was curious (as suggested by [94]), would improve understanding of the agent's internal state and have additional effects on participants.

In contrast to prior work, our robot expressed its curiosity through on-topic question-asking. In all three conditions, the robot exhibited general enthusiasm for the game, and made both correct and incorrect guesses with the same level of accuracy, to convey to the participant its role as a peer, rather than a knowledgeable teacher.

Curious+Reveal In this condition the robot was curious about rocks and the participant's curiosity about rocks, *and* it made statements that revealed why it was curious. The reason behind the robot's curiosity varied; it was either because of novelty (e.g., "*I have never seen shiny rocks before! What could make them so shiny? Do you not wonder that too?*") or expectation violation (e.g., "*Huh. I would have expected the rock to look different. I am curious though, do the holes form when gas bubbles get trapped when the lava cools? Do you have any idea?*"), both of which are known causes of curiosity [18].

Curious The Curious condition also had the robot exhibit curiosity about rocks and the participant's curiosity, but it did not reveal its rationale. For example, the robot said, "*I am curious. Do the holes form when gas bubbles get trapped when the lava cools? Do you have any idea?*".

Neutral In the Neutral condition, the robot did not express any curiosity for the rocks or the participant's curiosity. For example, the robot said, "*I believe holes can form when gas bubbles cool*".

In both curious conditions, in addition to asking the participant general rock formation questions (as described above), the robot asked more specific questions about the participant's rocks, e.g., "*What type of rock is your rock?*", or "*What other features does your rock have?*". In the Neutral condition, rather than asking these specific questions, the robot would share information about its own rock, i.e., "*My rock is metamorphic*". This was done to keep the amount of words said by the robot relatively consistent across conditions.

Robot speech was carefully designed to convey the same informational content about rocks to participants in all three conditions. The content used in the statements or questions made by the robot were taken directly from an introductory video about rock formation and classification that participants and robot were shown prior to playing the game. (See Tables 3.1 and 3.2 for an example dialogue).

3.2.4 Research Questions and Hypotheses

The main research questions guiding this study were:

Q1: Are the curious robots perceived to be more curious than the neutral one? If so, why?

Q2: Does curious question-asking behaviour in a robot produce emotional and/or behavioural curiosity contagion?

Q3: What impact does providing rationale for curiosity have?

Q4: Does curious robot behaviour affect learning?

Our hypothesis was that the robot's curious behaviour would elicit both behavioural and emotional curiosity contagion effects in participants, that could be about rocks and/or the robot. Specifically, participants would exhibit more of the curiosity-driven behaviour modelled by the robot, i.e., question-asking. Motivated by the elicited curiosity-driven behaviour, participants may pay more attention during the game and think more in depth about the content, thus learning more in the process, or actively seek information to gain more knowledge on rocks. Additionally, the robot revealing the rationale behind its curiosity would provide participants with a better understanding of the robot's internal model, resulting in more pronounced effects on learning, and curiosity contagion.

3.2.5 Methodology

Participants

30 students [20 female, 8 male, 2 other; age range: 18-49, mean 23, median 22] were recruited from a research-based university and randomly assigned to a condition. Participants volunteered for the study by responding to posters, and varied in their degree programs (i.e., Computer Science to Psychology) and level of education (7% PhD, 10% Masters, and 83% Bachelors students). Both native (63%) and non-native (37%) English speakers participated in the study.

Materials

The materials used in the game are shown in Figure 3.1a. In addition to the robot NAO, a small humanoid robot developed by SoftBank Robotics, there were 18 rocks, playing cards, and stands for the cards. There were 14 additional rocks which were used for the quizzes and free period (both described in more detail in Procedure).

Table 3.1: Illustration of LinkIt! dialogue, with examples of differences between conditions (P = Participant, R = Robot).

Neutral	Curious	Curious+Reveal
	P: I'll pick a new card. It's a feature card, glassy surface.	
	R: Who goes first?	
	P: I think number 4 has a glassy surface. Does it?	
	R: Alright, let me check my card. Yeah! That's correct! Hmmm, let's have a look at your rocks. Can you show me rock H?	
R: Some rocks have a really glassy surface. They probably cooled too quickly to form crystals.	**R**: I am curious why some rocks have a glassy surface. Could they have cooled too quickly to form any crystals? What do you think?	**R**: A glassy surface is new to me! I am curious why some rocks have that. Could they have cooled too quickly to form any crystals? What do you think?
	P: Um. I think you're right. So, if they cool down really fast they become glassy. Do you choose rock H?	
	R: Alright. I pick that rock.	
	P: Indeed, it has a glassy surface. You're right!	
	R: Yay! We linked the rocks!	
	P: Woohoo!	
R: My rock is igneous.	**R**: What rock type is your rock?	**R**: What rock type is your rock?
	P: It's igneous.	
	R: Cool! Mine too!	
R: I think all igneous rocks come from volcanoes.	**R**: I am curious whether all igneous rocks come from volcanoes. Do you think so?	**R**: I am curious whether all igneous rocks come from volcanoes. Do you think so?

Table 3.2: Table 3.1 Continued

Neutral	**Curious**	**Curious+Reveal**
	P: I think all igneous rocks come from lava or magma. So, I guess we could say they all come from volcanoes.	
R: My rock is shiny.	**R**: What other features does your rock have?	**R**: What other features does your rock have?
	P: It has multiple colours.	
	R: I see! They do look kind of similar.	
R: My rock is called Obsidian.	**R**: What's the name of your rock?	**R**: What's the name of your rock?
	P: Obsidian.	

Wizard of Oz Interface

The SoftBank Robotics Python SDK (version 2.1.4.13) was used to teleoperate the robot. An interface (programmed in Javascript, and deployed as a new application inside the robot, accessible at NAO's IP address) was implemented to allow for quick and easy selection from a set of predetermined statements and questions—supporting a more 'natural' interaction, as the possibility for long pauses was reduced, and consistency of robot responses between trials was ensured. Additionally, hand and arm gestures were evoked through the *ALAnimatedSpeech* module of the Python SDK, and the built-in "joyful" style was applied to NAO's voice to convey a positive attitude towards the game. The wizard handled participants speaking out of context by replying that it does not know about anything other than rocks.

Procedure

The procedure consisted of Session 1 (90 minutes) and Session 2 (a 30-minute session one week later). Participants interacted with the robot in Session 1; the procedure is shown in Figure 3.2. Both sessions were audio and video recorded.

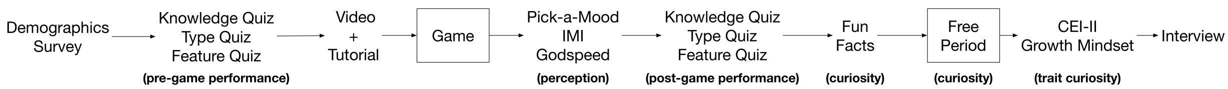


Figure 3.2: Study Procedure for Session 1 - with the variables being measured shown in brackets

Session 1—Pre-Game Session 1 began with the information letter, consent form, a demographics questionnaire, and three short quizzes (taking 10 minutes total) on rock classification, to provide a baseline from which learning gains could be measured. The *Knowledge quiz* involved multiple choice questions on rock formation. The *Type* and *Feature quizzes* had participants inspect the 18 rocks that would be used in the game (game rocks), plus 6 rocks not in the game (non-game rocks), and identify their types and features. After the quizzes, participant and robot together watched a 3-minute video describing rock formation and classification. They watched together so that the participant was aware of how the robot knew about rocks. Next, LinkIt! was explained to both the participant and robot, during which the robot asked clarification questions, introduced itself to the participant, and asked for their name, in order to calibrate the participant's expectations on the capabilities of the robot and how they could converse with it. After the explanation, participant and robot were told they would be left alone in the room for 30 minutes while they played the game.

Session 1—Post-Game Following the game, participants were given a set of questionnaires about their experience and their perception of the robot. First, was the "pick-a-mood" self-report scales [40]; one for themselves and one for the robot, asking them to select one (or more) out of 9 characters expressing emotion. Second, was the standardized Intrinsic Motivation Inventory (IMI), measuring their self-reported feeling of enjoyment, competence, effort, and relatedness with the robot [38, 157, 158]. Third, participants were given the Godspeed questionnaire [10], which consists of semantic differential scales on anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of the robot. After the questionnaires, participants proceeded to complete the post-game Knowledge, Type, and Feature quizzes, which were identical to the pre-game quizzes. In addition, to understand what rocks participants may be curious about, participants were given the opportunity to ask the researcher for a *fun fact* about any of the 24 rocks in the quizzes, with the constraint that the researcher would only tell them about a maximum of three of the chosen rocks.

Session 1—Free Period Next, participants were given 5-10 minutes to freely choose how to interact with the robot. During pilot studies, immediately after participants were told the study was over, they exhibited curiosity-driven behaviours, e.g., asking questions about the remaining rocks in the game, or about the robot. The free period allowed us to apply a novel procedure for assessing curiosity through behaviour in a *Free Choice Curiosity Test*, which was designed to enable systematic observation and quantifying of a participant's curiosity about rocks and the robot.

Participants were given eight rocks not used in the game or quizzes, and four 'ground truth' cards (with three features, the type, and the name listed on them) for four of the eight rocks. Participants were told they were free to do what they wanted with the rocks and the robot, and that the researcher would step out for 10 minutes. By presenting participants with eight new rocks, four with the ground truth, and four without, a measure of curiosity was constructed: If participants decided to ask the robot about any of the rocks for which they had the ground truth, it suggested that the participant was more curious about testing the robot's knowledge than knowing about the rocks. However, if they chose to ask the robot about the rocks for which they did not have the ground truth, it indicated curiosity for the rock.

If the robot was asked about any of the rocks, it would correctly identify the features and type, because, if the robot was incorrect for any of the four rocks for which participants had the ground truth, they may notice that the robot was incorrect and therefore, even though they are curious about the other rocks, not ask the robot about them. The robot's personality: Neutral, Curious, or Curious+Reveal, carried over from the game to the free period.

Session 1—Post-Free Period After the free period, participants filled out the PERTS Growth Mindset Assessment [123] and the Curiosity and Exploration Inventory-II (CEI-II) [78], both measuring *trait* curiosity–the characteristic of always having an interest in learning or obtaining new information. In contrast, *state* curiosity describes curiosity elicited by external situations—measured in our study by fun facts, verbal behaviour during the free period and game, and Likert scale questions on curiosity (described below).

Finally, a semi-structured interview was conducted with participants to understand their perception of the robot, the rationale behind their behaviour during the free period, as well as their overall experience. To measure participants' perception of the robot's curiosity, they were first asked whether they thought the robot had any particular personality. This was followed by separate Likert scale rating questions on how enthusiastic, engaged, and curious participants felt the robot was during the game. The questions were asked in this order so as not to make curiosity a focal point of the interview and minimize priming effects and bias. In general, the word *curiosity* was explicitly avoided in any questionnaires or interview questions until the end of the session. Participants were also asked whether *they* felt enthusiastic, engaged, and curious during the game.

Session 2 One week later, participants returned for Session 2. They were given the one-week Knowledge, Type, and Feature quizzes (which were the same as the pre- and post-game quizzes), and they could again ask the researcher for a fun fact about any of the 24 rocks in the quizzes. They were subsequently interviewed about whether they had thought about, or had any questions come to mind about rocks or the robot during the preceding week, and whether they had looked up any information related to either. We also wanted to know what they remembered from their time with the robot. Lastly, participants were debriefed on the study.

Table 3.3: Data Summary Table

Background	gender, age, degree, native_english_speaker
Perception	anthropomorphism, animacy, likeability, perceived_intelligence, perceived_safety, robot_mood, participant_mood
Trait	stretching, exploration, growth_mindset
Experience	enjoyment, competence, effort, relatedness
Curiosity	[pre|post|1wk]_curiosity_rocks, [pre|post|1wk]_curiosity_robot, [post|1wk]_funfact_rock[x]
Behaviour	avg lines, avg words, qa, ontask_qa, social_qa, uncertainty, argument, justificaiton, suggestion, agreement, disagreement, idea_verbalization, information_sharing, hypothesis_generation
Learning	[pre|post|1wk]_quiz_accuracy, [pre|post|1wk]_type_accuracy, [pre|post|1wk]_feature_[precision|recall|accuracy|specificity]

3.3 Analysis

3.3.1 Data Preparation

There were several sources of ambiguity in the quizzes as a result of being filled out on paper. In the Feature quizzes, participants on a few rare occasions marked both a feature and 'I don't know'. In this case, the marked feature was taken as the final answer. When participants marked both a feature and 'None of the Above', the answer was discarded. The data of one participant (C8) in the Curiosity condition was discarded entirely, since it was unclear whether the participant understood the procedure of the experiment. In total, the data of 29 participants was retained for analysis.

The audio of the game was transcribed and using the theoretical framework of curiosity from Sinha, Bai, and Cassell [145], ten verbal curiosity-driven behaviours were coded: (1) *Uncertainty:* e.g., "I'm not sure, but I think that is why it happens", (2) *Argument:* e.g., "Ok we can pick a new card but I'm gonna tell you you're wrong because rock G, it has the feature cemented together", (3) *Justification:* e.g., Robot: "How do you know the name of the rock is conglomerate?" Participant: "Um there's little pebbles in it. And the little pebbles are like a conglomerate of all the things put together", (4) *Suggestion:* e.g., "Here, choose one rock and tell me why you like that rock", (5) *Agreement*, (6) *Disagreement*, (7) *Question Asking:* (on task vs. social) e.g., on task - "What other features does your rock have?", social - "What time did you wake up this morning?", (8) *Idea Verbalization:* e.g., Robot: "I wonder why some rocks have so many layers. Do you know why?" Participant: "Because they are usually... they get formed over time", (9) *Sharing Information:* In the game interaction this described the voluntary sharing of information about a rock, e.g., "It's magnetic too!", and (10) *Hypothesis Generation:* e.g., "So I'm not really sure. I guess it's just from being made out of different minerals". Each statement/question could be associated with multiple codes.

The Free Choice Curiosity Test was designed to measure whether elicited curiosity was directed towards the robot or the rocks. The audio of the free period was transcribed, and both the audio and video were coded with a rank assigned to each rock to indicate the order in which participants chose to talk about them. The premise was that participants would choose the rocks that they were most interested in first.

3.3.2 Measures

A variety of data were collected about participants' demographics, curiosity, perception of and experience with the robot, behaviour in the game and free period, and learning. The focus is on those measures that provide, primarily, insights into participants' mental state of curiosity and curiosity-related behaviour, and secondarily, into learning.

The dependent variables include demographics information, perception of the robot (e.g., Godspeed measures, pick-a-mood scale), trait curiosity (e.g., curiosity and exploration inventory, growth mindset survey) as well as intrinsic motivation. Table 3.3 briefly summarizes all the variables considered in the analysis along these dimensions. Only a few of the variables are found to be significant through the step-wise selection process; these variables and their associated results are reported in the findings.

3.4 Results

3.4.1 Perception of Agent Curiosity

During the Session 1 interview, participants were asked to rate on a 7-point Likert scale how curious they thought the robot was from 1 (not at all curious) to 7 (extremely curious). Table 3.4 shows the average rating for each condition. Participants in the Curious and Curious+Reveal conditions perceived the robot to be more curious than participants in the Neutral condition. The proportional odds model [2] showed that the robot's curiosity was ranked significantly differently by the participants in the different conditions ($\chi^2(2, N = 28) = 7.46, p = 0.02$). This demonstrates that our design created a robot that behaves in a manner considered curious by participants.

Table 3.4: Perception of Robot Curiosity about Rocks

Condition	Average Rating (SD)
Neutral	3.78 (2.33)
Curious	5.89 (0.78)
Curious+Reveal	5.90 (0.97)

Approximately half of the participants in the curious conditions, and one participant in the Neutral condition, described the robot as 'curious', 'inquisitive', or having a desire to learn—before being asked specifically by the interviewer about the robot's curiosity. The majority of participants in the Curious (C) and Curious+Reveal (CR) conditions stated that it was the question-asking behaviour which indicated to them that the robot was curious: "The robot actually actively asked me for more knowledge" (C1), and, "He's very curious because he kept asking a lot of questions" (CR6). The curious robots were perceived to be curious not only about rocks, but also about the participants' state of mind. Participant C2 said, "The robot seemed curious about my opinions on things. So he would make a connection and say 'What do you think?"'. Participants also noticed a link between asking questions and a desire for new knowledge: "[The robot] wants to know more" (C7); "He kind of always wanted to learn about, and go a step further to learn a little bit more about the rock. He was kind of trying to go deeper" (CR5).

Rather than asking questions, participant C5 found the robot to be curious because "when we were playing the game, I was moving to the next card, and he [was] still talking about the rock". Curiosity was also perceived from the fact that "he always notices little details about [the rocks], and connects it to what he knows and tries to share with me" (CR2).

In contrast, the participant who rated the robot's curiosity the lowest in the Curious+Reveal condition, felt that "if NAO were to be curious, then it would ask me different kinds of questions every time" (CR4). The participant was referring to the robot asking about the participant's rocks features, type, or name, after every time the rocks linked: "I felt that if you ask the same thing again and again...it's more robotic".

Unexpectedly, when asked: "Do you think [the robot] was curious or not curious about rocks during the game?", 6 out of the 10 participants in the Neutral (N) condition felt that it was curious. However, on average, they gave a significantly lower Likert scale rating of the robot's curiosity (Table 3.4), implying that participants in this condition did not think it was as curious as participants who interacted with the curious robots. In the absence of the robot asking questions to know more about rocks, participants picked up on cues that were not intended to convey curiosity: "He was curious because he *knew* a lot about [the rocks]" (N1); "NAO is really *interested* in playing the game... really pushing the game to get it going" (N2); and "He would tell me random facts" (N3). Non-verbal cues of curiosity were also noticed by some participants, i.e., "He tilts his head" (N3), and, "I think he's actively looking for certain traits when you give him a rock" (N7).

Additionally, two participants mentioned that the robot asking questions related to game play indicated curiosity to them, e.g.: "I think he was curious about what I would do, like [he would say] 'Do you want to go first?"' (N3). The remaining participants in the Neutral condition found that the robot was not curious specifically because "he didn't ask me questions" (N4). Participant N9 explained: "To me curiosity comes from wanting to know something....So the way he sounds is curious, but I think that's from the enthusiasm but in terms of actual curiosity maybe not". The observation about question-asking by some participants in the Neutral condition, further supports that our design of verbal question-asking behaviour by the robot can be used to convey curiosity.

3.4.2 Participants' Curiosity

About Rocks

The demographics questionnaire (pre-study) asked participants to rate their level of interest in rocks on a 7-point Likert scale from 1 (not at all) to 7 (very). 17% of participants were not at all interested in rocks, 73% were interested, and 10% were very interested. During the interviews in Session 1 (post-interaction) and Session 2 (one-week), in addition to rating the robot's curiosity, participants were asked to rate their *own* curiosity about rocks, also on a 7-point Likert Scale, from 1 (not at all curious) to 7 (extremely curious). Table 3.5 shows the average rating in each condition at pre-study, post-interaction, and after one-week. The differences in pre-study curiosity on rocks between the conditions were significant. This bias was controlled for by comparing the magnitude of change in curiosity across the three time points, between conditions. Compared to the other conditions, participants in the Curious condition experienced a greater change (from pre-study) in curiosity about rocks after the game ($F(2, 26) = 4.12, p = 0.03$) and after one week ($F(2, 26) = 5.47, p = 0.01$).

Table 3.5: Average Participant Curiosity about Rocks (SD)

Condition	Pre-Study	Post-Interaction	One-Week
Neutral	4.00 (1.05)	4.90 (1.35)	4.90 (1.45)
Curious	3.11 (1.05)	4.61 (1.32)	3.67 (1.00)
Curious+Reveal	4.50 (1.27)	4.50 (1.43)	3.75 (1.55)

For the requested fun facts, a Poisson regression model was used with Neutral as the baseline category, to which Curious and Curious+Reveal were compared. Participants in the Curious ($\hat{\beta} = 2.09, t(364) = 2.51, p = 0.01$) and Curious+Reveal ($\hat{\beta} = 1.21, t(364) = 1.96, p = 0.05$) conditions tended to ask to learn more fun facts about the rocks the robot had commented on during the game. After one week, the conditions did not appear to be significantly different anymore, $\hat{\beta} = 1.45, t(364) = 1.55, p = 0.12$ (Curious) and $\hat{\beta} = -0.46, t(364) = -0.67, p = 0.50$ (Curious+Reveal). However, there is evidence that the more questions asked by the robot during the game about the participant's rocks type, the lower the probability that a participant would inquire for a fun fact about that particular rock a week later, $\hat{\beta} = -0.22, t(364) = -2.06, p = 0.04$.

Analysis of the Free Choice Curiosity Test, under the Kruskal-Wallis test, showed that the rank order for all rocks was similar in all conditions, suggesting that participants did not prefer the rocks with the ground truth over those without, and vice versa. However, in addition to coding rank order, participants were asked to explain in the interview what they did during the free period and why. Only participants in the curious conditions described their behaviour as being solely about rocks (five participants in the Curious condition and five in the Curious+Reveal condition). Common explanations for behaviour were: using the robot to clarify or get advice on distinguishing rock types, to get more knowledge on rocks, or using the robot to test whether their own knowledge was correct. All participants in the Neutral condition stated their behaviour was either directed solely towards the robot, or both the rocks and the robot.

In Session 2, participants were asked in the interview whether they had had any questions or had thought about rocks at all since Session 1, and whether they had looked up any information on the topic. Seven participants in the Curious condition and six in the Curious+Reveal condition, compared to only three in the Neutral condition, stated that they had questions about rocks, and three participants in each condition explained they had actively searched for information about rocks. Participants in the Neutral and Curious+Reveal conditions who said they looked up more information, were mostly interested in clarifying their understanding between sedimentary and metamorphic rocks, because they were unsure of how to categorize those types. In contrast, participants in the Curious condition stated that they searched for information because, “it’s interesting” (C4), and, “I was just thinking about it and just wonder, I’m like more intrigued basically, because of the process” (C6).

About the Robot

Similar to ranking their curiosity about rocks, participants were asked in the demographics questionnaire (pre-study) to rate their interest in robots on a 7-point Likert scale from 1 (not at all) to 7 (very). 3% of participants were not at all interested in robots, 67% were interested, and 30% rated themselves very interested. In the interview during Session 1 (post-interaction) and Session 2 (one-week), participants were asked to rate their level of curiosity about the robot on a 7-point Likert scale from 1 (not at all curious) to 7 (extremely curious). The average ratings per condition at each of the three time points is shown in Table 3.6. In contrast to curiosity about rocks, participants’ curiosity about the robot, in all conditions, started relatively high, increased immediately after the game, and decreased one week later. However, these changes are not statistically different between conditions after the game ($F(2,24) = 0.84, p = 0.44$) and one week later ($F(2,21) = 0.23, p = 0.79$).

Table 3.6: Average Participant Curiosity about Robot (SD)

Condition	Pre-Study	Post-Interaction	One-Week
Neutral	4.90 (1.60)	6.11 (0.78)	5.60 (1.26)
Curious	4.89 (0.60)	6.50 (0.53)	5.36 (1.03)
Curious+Reveal	5.30 (0.95)	5.80 (1.81)	5.50 (1.32)

All but one of the participants rated their curiosity about the robot equal to, or higher than, their curiosity about rocks post-interaction. The participant who was not curious about the robot stated, “Well it’s just a computerized robot, right?” (CR7). The other participants were mainly curious about how the robot worked, what else it could do, what else the robot knew, and whether it was learning. For example: "I'm pretty curious about NAO... [If I] phrased things differently or asked a question in a non-direct way...I couldn't stump her. That was cool" (CR8), and, "I was very curious about how he is identifying [the rocks]" (N10), but also, "I was curious about whether he'll build on, you know, different experiences. So after my trial, would it differ from someone else's?" (CR6). Participant C9 made an interesting observation, stating: "I think I'm a lot more interested in NAO. Like he has a *contagious* influence".

When asked to explain what they chose to do during the free period, only one participant in the Curious condition stated they were only focused on the robot. They explained: "I kind of tested him...how this robot can...like what's his intelligence [and] I specifically used the rocks with the cards to kind of see if he can correctly, precisely identify features" (C3). Across all conditions, when participants showed some curiosity towards the robot, most explained how they tested it to figure out what more it could do and what more it knew. Participant CR3 taught the robot the names of the rocks and then tested to see whether it could remember them. Participant N1 wanted to see whether the robot could distinguish between numbers and letters that looked similar. Another stated, "I wanted to know if [the robot] could do anything wrong. I wanted to challenge him... I wanted to see if he was gonna get mad" (N3).

In Session 2, the average rating for curiosity about the robot was approximately the same in each condition. Six people in each of the three conditions said they still had some questions pertaining to the robot, however very few searched for more information.

3.4.3 Participants' Verbal Behaviour

Participants' verbal behaviour was coded using the theoretical framework of curiosity [145]. As a validity test, two independent coders each coded half of the game transcript dialogues. On average, the inter-rater agreement (cohen kappa score) was around 0.74, indicating substantial agreement. The lowest agreement rate was 0.50 (i.e., moderate agreement) for Justification, and the highest was 0.85 (i.e., almost perfect agreement) for Question Asking.

Table 3.7 shows how participants' responses differed by condition, with the statistically significant response types in bold. Our findings, using ANOVA, show significant differences in the frequency of certain types of participant responses between conditions. Participants in the curious conditions produced a greater number of responses in the game that can be categorized as expressing curiosity, compared to the Neutral condition. In particular, there was more question-asking, specifically on-task question-asking ($F(2, 26) = 7.27, p = 0.003$), in the Curious and Curious+Reveal conditions compared to the Neutral condition. There are also statistically significant differences in responses expressing uncertainty ($F(2, 26) = 18.15, p < 0.001$), hypothesis generation ($F(2, 26) = 13.65, p < 0.001$), and disagreement ($F(2, 26) = 6.72, p = 0.004$). Across all conditions, there was very high verbal agreement with the robot. This is most likely a consequence of the participants not having enough knowledge on rocks to disagree Disagreement with the robot, although low, was significantly higher in the Curious+Reveal condition—possibly a result of the robot providing rationale, and giving participants more explanation with which they could disagree.

Table 3.7: Average Number of Participant Responses by Type in Game (SD)

Condition	Neutral	Curious	Curious+Reveal	p
Question Asking (QA)	1.30 (1.34)	9.67 (7.26)	3.40 (2.37)	***
On-Task QA	1.00 (1.25)	8.89 (7.79)	3.20 (2.49)	**
Social QA	0.30 (0.48)	0.78 (1.99)	0.20 (0.63)	
Uncertainty	0.40 (0.70)	8.44 (3.50)	6.00 (3.83)	***
Argument	0.00 (0.00)	0.00 (0.00)	0.10 (0.32)	
Justification	0.40 (0.52)	2.33 (2.24)	2.00 (2.16)	.
Suggestion	0.00 (0.00)	0.11 (0.33)	0.00 (0.00)	
Agreement	9.60 (4.93)	9.67 (2.12)	10.5 (4.53)	
Disagreement	0.00 (0.00)	0.67 (0.71)	1.00 (0.82)	**
Idea Verbalization	0.00 (0.00)	0.11 (0.33)	0.00 (0.00)	
Information Sharing	5.00 (4.35)	4.00 (3.84)	1.50 (1.65)	.
Hypothesis Generation	0.00 (0.00)	3.78 (1.30)	1.90 (2.38)	***

3.4.4 Other Measures

Learning

For the Knowledge and Type quizzes, the score is 1 if the answer was correct for a particular question/rock, zero otherwise. For the Feature quiz, as the answer could include a number of features, 5 different measures were computed—precision, recall, F1, specificity, and accuracy, to capture both how many features participants correctly identified and how many features they missed. Condition had no effect on the Knowledge, Type, and Feature quiz scores, i.e., learning; nor did it have an effect one week later, i.e., retention. However, through step-wise linear regression, participants who reported being more curious about rocks after the game were found to have higher Knowledge quiz scores a week later, $\hat{\beta} = 0.32, t(26) = 2.34, p = 0.03$.

Results show that participants performed much better on determining the type of rocks they saw in the game than the rocks not in the game, as reflected in the Type quiz scores immediately after the game ($F(2, 83) = 22.64, p < 0.001$) and one week later ($F(2, 83) = 20.69, p < 0.001$). Similarly, there was significantly different performance for game rocks versus non-game rocks for both the post and one-week Feature identification quiz. These results show that the LinkIt! game has some success in teaching participants how to identify features in rocks and determining the rock type.

To investigate the measures that impact the number of questions participants answered correctly, different Binomial regression models were used for the post and one-week Knowledge quiz, respectively. Through step-wise selection, the resulting models show that participants' initial number of correct answers significantly influence the number of correct answers in the post Knowledge quiz ($\hat{\beta} = -0.37, t(27) = 2.24, p = 0.03$) but not the one-week Knowledge quiz ($\hat{\beta} = -0.13, t(25) = -0.27, p = 0.79$).

Perception of the Robot

General There were no significant differences between condition in the four Godspeed measures: anthropomorphism, animacy, likeability and perceived intelligence.

Gender The robot's voice was child-like, without any explicit manipulation of pitch to indicate gender, and the researcher conducting the sessions referred to the robot with the pronoun 'he'. However, other cues appeared to influence perception of the robot's gender, as all participants in the Neutral condition used 'he', but two participants in the Curious condition and three in the Curious+Reveal condition, referred to the robot as 'she', while the rest used 'he' or 'it/the robot'.

Intrinsic Motivation Inventory

No statistical significance was found between different conditions in terms of participants' self reported measures of enjoyment, competence, effort and relatedness on the Intrinsic Motivation Inventory.

3.5 Discussion

This study presents a novel structured game, LinkIt!, for the design and study of robot behaviour and human-robot interaction, and introduces a procedure for assessing curiosity through behaviour, called the Free Choice Curiosity Test. The findings show that verbally-expressed curiosity through on-task, topic-directed question-asking, can reliably be recognized as curiosity in a social peer robot, and provide evidence for curiosity contagion effects; i.e., the robot's verbal expression of curiosity about rocks increased the participants' emotional *and* behavioural curiosity about rocks. Participants in the Curious condition reported a greater increase in curiosity about rocks after the game, and participants in the Curious and Curious+Reveal conditions tended to ask to learn more fun facts about rocks that the robot had commented on. In Session 2, more participants in the curious conditions than in the Neutral condition had questions about rocks and were interested in gathering new information about rocks. Furthermore, results from the Free Choice Curiosity Test indicate that the curious robots were more effective at directing participants' curiosity towards the rocks than the Neutral robot.

As with Gordon et al.'s findings [56], the results show behavioural curiosity contagion effects as a result of interacting with a curious robot. Different from Gordon et al., this study focused on the verbal behaviour of participants interacting with curious robots, rather than curiosity-driven exploration and uncertainty seeking. Using the theoretical framework of curiosity [145] to code dialogue, a significant increase in on-task question-asking by participants in the Curious condition was found. This result could indicate the *chameleon effect* [32], the unconscious mimicry of behaviour–a well-documented phenomenon in human social interactions. The study shows that within a 30 minute interaction, students appear to mimic the most pronounced verbal behaviour of a robot dialogue partner. In addition, beyond on-task question-asking, participants in the Curious condition also more frequently expressed other curiosity-driven behaviours such as uncertainty and hypothesis generation in their conversation with the robot, indicating emotional curiosity contagion effects as well as behavioural.

The findings also show significant differences in the frequency of certain types of participant responses between conditions, which can be the result of the dynamics of the conversation itself. For example, question-asking by the curious robots may have provided more opportunities for participants to generate hypotheses, than the neutral robot that only shared information. Detailed conversational analysis is left to future work, however it is of note that verbal mimicry of question-asking did not occur significantly more in the Curious+Reveal condition, even though this robot also asked questions. Furthermore, participants in this condition did not rate their curiosity about rocks higher after playing the game. Tsai et al. [154] propose that emotional contagion effects may be hindered by cognitive workload. The addition of rationale in the Curious+Reveal condition may have been too informationally overwhelming for participants, leading to neither behavioural *nor* emotional curiosity contagion—indicating that the mimicking of question-asking behaviour by participants in the Curious condition may have aided in emotional curiosity contagion.

Similar to prior work [56], the study found that the manipulations had no effect on learning; neither immediately after the interaction nor one week later. This remains an interesting finding as our study considered a different age group, interaction protocol, and expression of curiosity. The result may be due to the novelty effect—an increase in performance when interacting with a new technology, as a result of *interest* in the new technology rather than an actual improvement in learning. For all participants, interacting with a social humanoid robot was a novel experience, and many expressed surprise and curiosity about the robot. In the future, this effect could be mitigated by having participants interact with the robot over multiple sessions and measuring their overall learning performance. However, it was found that participants who reported being more curious about rocks after the interaction performed better on the Knowledge quiz one week later, indicating that curiosity elicited in Session 1 may have helped retain knowledge for Session 2.

A few participants noticed themselves become curious about rocks as a consequence of learning more about them. This supports the information gap theory of curiosity [96], which postulates that curiosity originates from a gap between what one knows and what one wants to know. Once people become aware of this gap, their curiosity is piqued. It appears that the question-asking by the curious robots acted to bring attention to the information gap and make participants aware of its existence, and in this way elicited curiosity. It also supports the Learning Progress hypothesis [118], which suggests that the experience of learning causally influences curiosity and an intrinsic motivation for new knowledge.

The robot in this study was naive; it only knew about what it saw in the introductory video and what it learned from the game. As a result, even when participants became aware of their own knowledge gaps and were asking the robot for answers, the robot was not always able to provide them. Previous research has shown that when people do not expect to close their knowledge-gap quickly, not knowing can affect their subjective experience of curiosity [112]. Similarly, Shiomi et al. [142] found that students were motivated to ask more questions when they *knew* the robot could answer them. In future studies, giving the robot the ability to answer participants' questions, or providing participants with material they could search to resolve their information gaps, may lead to further increases in curiosity.

The novel Free Choice Curiosity Test did not show significant differences in how participants chose rocks to present to the robot. Instead, insights were gained into participants' curiosity from their explanations—indicating a need for development of how the test is conducted, as well as analyzed, in order for it to be an effective objective measure of curiosity.

3.5.1 Limitations

A limitation of this study is the relatively small sample size in each condition and the fact that all participants are university students. This makes it difficult to generalize the results to larger or different populations. To handle small sample size, the analysis avoided including more covariates in the ANOVA and selected the simplest possible model to explain our data in order to prevent overfitting, while providing some general directions for future research involving a larger sample.

Both small sample size and large individual differences may have contributed to the lack of significant differences between conditions in certain measures. In the analysis, individual differences were considered by using curiosity trait measures as independent variables, however, these were not found to be significant. Future work could run the study with younger students and include more participants. A within-subject experimental design may be adopted to combat large individual differences, and other traits could be measured that are known to influence learning and social interactions, such as executive, emotional, and social functioning, as well as learning goals (e.g., [140, 95]).

Furthermore, non-verbal indicators of curiosity during the game were not coded. In the future, video could be coded for such indicators to provide a more complete picture of emotional and behavioural curiosity contagion. For example, one possibility is to use facial landmark detection to analyze facial action units to infer affective states that often occur with curiosity as Sinha, Bai, and Cassell [145] have done with human-human social interaction.

3.5.2 Design Implications

The findings from this study indicate that care must be taken in the design of robot dialogue to prevent high cognitive workload in the conversational partner. Furthermore, curiosity can be inadvertently communicated through cues other than question-asking (e.g., body language, task-relevant behaviours, etc.). Lastly, more research can be done on behavioural measures of curiosity, such as the Free Choice Curiosity Test.

3.6 Conclusion

This study was designed to investigate whether a pedagogical agent can convey curiosity (a form of intrinsic motivation) to learners through on-topic question-asking, a verbal behavior that had not yet been explicitly implemented as a means of expressing curiosity, and whether this expression could affect learners' own curiosity, as well as enhance the cognitive outcome of recall. The results suggest that on-topic question asking creates an agent that behaves in a manner considered curious by learners, it elicits both emotional and behavioural indicators of motivation, but it has no effect on the cognitive learning outcome. Furthermore, if the agent provides too much information about it's curiosity (as seen in the Curious+Reveal condition), it's possible that emotional and/or behavioural contagion effects are hindered by the extra cognitive workload imposed to process the additional information.

Chapter 4

Humour: Creating a positive learning experience

4.1 Related Work

Pedagogical agent researchers have suggested that a positive learning experience can influence learning outcomes by providing cognitive rest, increasing engagement, and evoking memory cues (e.g., [62]). Various features of pedagogical agents have been explored in order to facilitate a positive learning experience (e.g., reassuring learners [63]), and based on Psychology literature, in education, humour is proposed to lead to a more positive emotional and social environment.

Humour is an integral part of human communication and can be expressed verbally (produced by means of language or text, e.g., jokes, comics) and/or non-verbally (e.g., facial expression, gesture), intended to elicit responses such as laughter and mirth [100]. Research demonstrates that material taught with humour can lead to better retention and recall [34, 29, 172, 50] and humour can increase student interest and ability to engage in divergent thinking [41, 172, 171]. The beneficial effects of humour may be the result of humour facilitating a more relaxed environment, allowing for cognitive rest and absorption of the information [79]. The benefits of humour in education have long been alluded to as a result of humour's psychological effects on learners, however, little research exists on the effects of humour in pedagogical agent contexts.

In the past it has been common for classroom humour to be used spontaneously rather than intentionally to achieve certain learning goals [29]. Berk [16] investigated and evaluated 10 strategies for using humour as a pedagogical tool in the classroom, specifically for reducing anxiety, improving students' ability to learn, and improving performance on exams and assignments. The 10 strategies were: humorous material on syallabi, humorous material on handouts, opening jokes, skits/dramatizations, in-class spontaneous humor, in-class spontaneous questions, in-class humorous examples, humorous problem sets, Jeopardy-style review of material, and humorous material on exams. He recruited 316 undergraduate and graduate students from introductory statistics courses. Overall, both graduate and undergraduate students found the strategies as being either Very Effective or Extremely Effective at *reducing anxiety*. Similarly, both graduate and undergraduate students rated all strategies as being either Very Effective or Extremely Effective at improving ability to learn. The 1995 and 1996 undergraduate classes rated in-class spontaneous humor, humorous examples, and problem sets as Extremely Effective, and opening jokes and skits/dramatizations as Very Effective all three years. For undergraduate students, only spontaneous humor was rated Very Effective for improving performance, all three years (1994, 95, and 96).

In the field of personality research, humour as a part of an individual's character and the role it plays on the way they engage with others, has generated growing interest. Such work has lead to distinctions between individual differences in humour style—behavioural tendencies related to the uses or functions of humour in everyday life. One of the most prominent contributions in the field distinguishes between four humour styles: *affiliative* (use of benign humour to enhance social relationships), *self-defeating* (use of humour to enhance relationships at one's own expense), *self-enhancing* (use of humour to enhance oneself), and *aggressive* (use of humour to enhance oneself at the expense of others) [101]. Affiliative and self-enhancing humour have been found to positively correlate with positive affect whereas, self-defeating humour has been shown to be negatively related to positive affect (e.g., [28]).

The style of humour used can impact the way in which a teacher is perceived. Ziv, Gorenstein, and Moris [173] found that students responded differently to a human teacher who used four different types of humour during a lecture: when using a combination of self- and other-disparaging (also known as self-defeating and aggressive) humour, the teacher was rated most appealing and original, when using only other-disparaging humour, the teacher was rated most powerful, and when the teacher did not use humour they were evaluated as having the most systematic teaching style. The study also showed that students who possess a sense of humour are most appreciative of a teacher using humour. Tamborini and Zillman [149] found no difference in rated intelligence when a college lecturer used sexual, other or self-disparaging humour. However, use of self-disparaging humour was perceived as influencing 'appeal'. Gruner [61] showed that speakers using self-disparaging humour were perceived as wittier than those not using humour.

Humour in Conversational Agents

Research on humour in conversational agents has also found the style and form to impact perception and interaction in various ways. In robots, four forms of humour (wit and an ice-breaker, corny jokes, subtle humour, and dry humour and self-deprecation) are suggested to enhance sociality of a robot [76], and innocent humour (riddles and punning jokes) was found to improve perception of task enjoyment, robot personality, and speaking style [110]. In interactions with virtual agents, conversational and situation-specific jokes have been found to affect how cooperation is perceived in an agent [83], humour is proposed as a means of recovering from error situations while providing a pleasant user experience [111], and affiliative humour has been shown to significantly motivate healthy behaviours [117].

Morkes, Kernal, and Nass [107] found that when participants communicated with a chatbot while working together on a task, but thought they were interacting with another human, the use of humour in on-task conversation lead participants to rate their conversation partner as more likable and reported greater cooperation with, and similarity to, their partner. They also made more jokes and responded more socially. On the other hand, when participants knew they were conversing with a computer agent, the use of humour lead to participants being less social, smiling and laughing less, and feeling less similar to their conversation partner. They also spent less time on the task. Conversely, Dybala et al. [44] reported that participants found a chatbot using puns in off-task conversation to be more natural, interesting, and easier to talk to than a non-humorous agent. In a follow-up study, the researchers also found that the humorous agent was rated as more human-like, funny, and likeable than the non-humorous agent [46]. The researchers describe two major subclasses of conversational agents that use humour: on-task and off-task. They state: "we believe that the presence of humour is of higher importance in non-task-oriented agents, for their main purpose is to entertain human interlocutors and socialize with them during the conversation". Thus, it appears that users are sensitive to an agent's humour, but the impact on interactive behaviour varies depending on the nature of the task.

Previous studies have highlighted the usefulness of pedagogical agents that have socially-oriented conversations with students, e.g., reassuring them [63], initiating small-talk [62], and engaging in mutual self-disclosure [146], leading to more positive experiences and promoting learning gains. Little published research however exists on using humour with conversational agents in learning contexts. One preliminary study [162] had a virtual human tutor use humorous jokes and pictures in an e-learning interface. The researchers found the humour to enhance learner motivation, performance, and ease emotions. These initial results point to the possibility of using humour as off-task social conversation with pedagogical agents to create a positive experience, and promoting learner motivation.

4.2 Study II

Although prior work indicates the value of humour expression by pedagogical agents, there is little published research systematically investigating the use of humour styles, especially in agents taking on the role of tutee, i.e., teachable agents. Taking into account the suggestions made by researchers for creating a positive learning experience, and humour having been shown to decrease stress and anxiety [54], a study was conducted to examine humour in a teachable agent context. The type of humour manipulated in the study was *humour style*. In particular, humour styles provide a form of affective expression for the agent, as they have been found to be associated with various moods; two humour styles were compared: *affiliative humour* – positively associated with such things as self-esteem, well-being, and positive moods, and *self-defeating humour* – negatively correlated with such measures.

4.2.1 Agent

A text-based agent was chosen for this study, for the following reasons. First, although the physical presence of robots has been found to produce cognitive learning gains as well as more positive interactions compared to voice-only or virtual agents [8, 92], they can be costly and inaccessible to learners. Second, although there is a strong connection between humour and laughter, laughter can exist separate from humour. For example, 'social/conversational' laughter can be distinguished from 'hilarious' laughter, the latter being more directly associated with humour [42]. Furthermore, not only are there various types of laughter, but it also provides social bonding signals and is "one of the most important non-verbal vocal social signals" [42], and researchers have investigated what the different characteristics of laughter are and how they affect the way it is perceived and its impact on listeners. Trouvain and Schroder [153] also suggest that laughter is a very specific and distinct type of auditory gesture that requires careful design/modelling for an effect on social bonding. As humour and laughter have not been extensively explored in pedagogical agent scenarios, this exploratory study focuses on the effect of verbal humour as expressed through humour styles in a text-based agent, and remove possible confounds from the addition of auditory laughter. It does therefore remain a lesser-explored area of humour in pedagogical agents, and presents an interesting opportunity for future research. Lastly, using a text-based agent allowed the study to be conducted online, meaning that learners could participate from more real-world settings, rather than in-lab, and it could continue during the COVID-19 pandemic.

4.2.2 Task

Using the *Curiosity Notebook* [85], participants taught a virtual text-based agent, Sigma, how to classify rocks. The gender of Sigma was never specified nor implied throughout the entire interaction, and Sigma was represented by a static avatar. The original interface and participant-agent dialogue was adapted and customized to fit this study's research questions (For reference, the original layout and task format are described in [85]). The teaching interface of the Curiosity Notebook (Figure 4.2) consists of a reading panel on the left-hand side, with a number of articles and pictures about the topic to be taught divided into different categories. Each rock is given its own article. The sentences in the articles could be selected and taught to the agent at certain moments during the interaction. A chat window, through which participants could converse with the agent, was on the right-hand side of the screen. Everything taught to the agent is recorded in a 'notebook' that can be viewed at any time (See Figure 4.1). Every rock is given a page in the notebook and the notebook updates live.

The text in the articles is adapted from https://geology.com. The sentences in each article are 'linked' to features necessary for classification (large or small crystals, layers, a glassy appearance, holes, sand or pebbles, fossils, and formation process). These linked sentences are used as a ground truth to verify that participants select the correct sentence to teach the agent. If the sentence selected does not match the feature or rock the agent asked about, the agent asks the participant to select another sentence that would better answer the question. It also uses this ground truth in answering quiz questions (described in more detail below).

Interaction

When the participant is ready to teach, they select one of seven buttons split into *Teach* (Describe, Explain, or Compare), *Check* (Correct or Quiz) and *Entertain* (Fun Fact or Tell Joke) groupings, which initiates the agent to begin a conversation. Although participants can choose the type of interaction to have with the agent, the agent drives the conversation by asking questions and making statements.

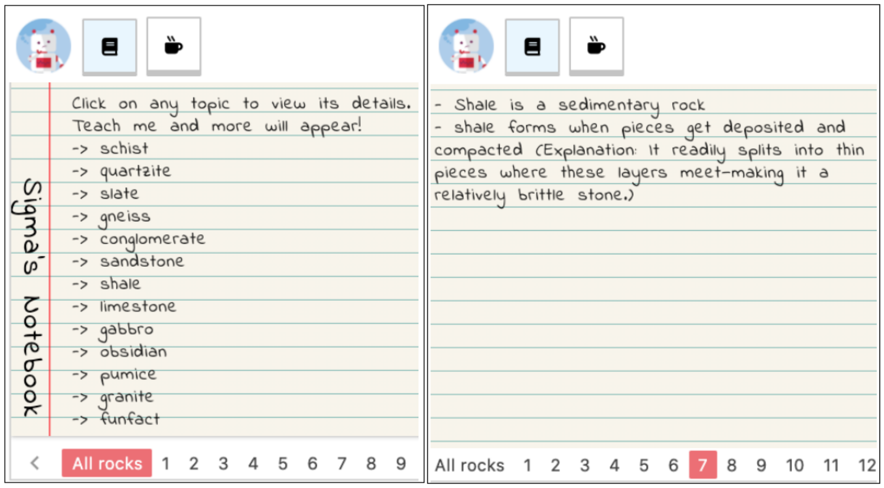


Figure 4.1: Sigma's Notebook, showing list of rocks taught so far (top) and page of notes on Shale rock, with an explanation provided by a participant (bottom).

Describe involves the agent asking for the name of a rock, what rock type it belongs to, and selecting a sentence from the articles with information on what feature(s) or characteristic(s) help classify it. In the Explain interaction, the agent asks why a rock is of a certain type or why it has certain features. Participants answer by selecting a sentence and are sometimes prompted for an explanation in their own words. To select a sentence, the participant is required to navigate to the article of the appropriate rock and find the sentence that contains the answer. Compare has the participant compare two rocks to each other - focusing on what is similar or different depending on whether the two rocks chosen are from the same category or different. The Correct button allows participants to change information they had taught the agent. The Quiz button gives participants the opportunity to ask the agent to classify a rock to assess its knowledge state. The Fun Fact button has the agent ask the participant to provide a fun fact and occasionally ask for an explanation of why they thought it was interesting, and the Tell Joke button allows participants to tell the agent a joke. Once the dialogue associated with a button finishes, participants can select another button, allowing them to take a break if needed, and decide what interaction to do next.

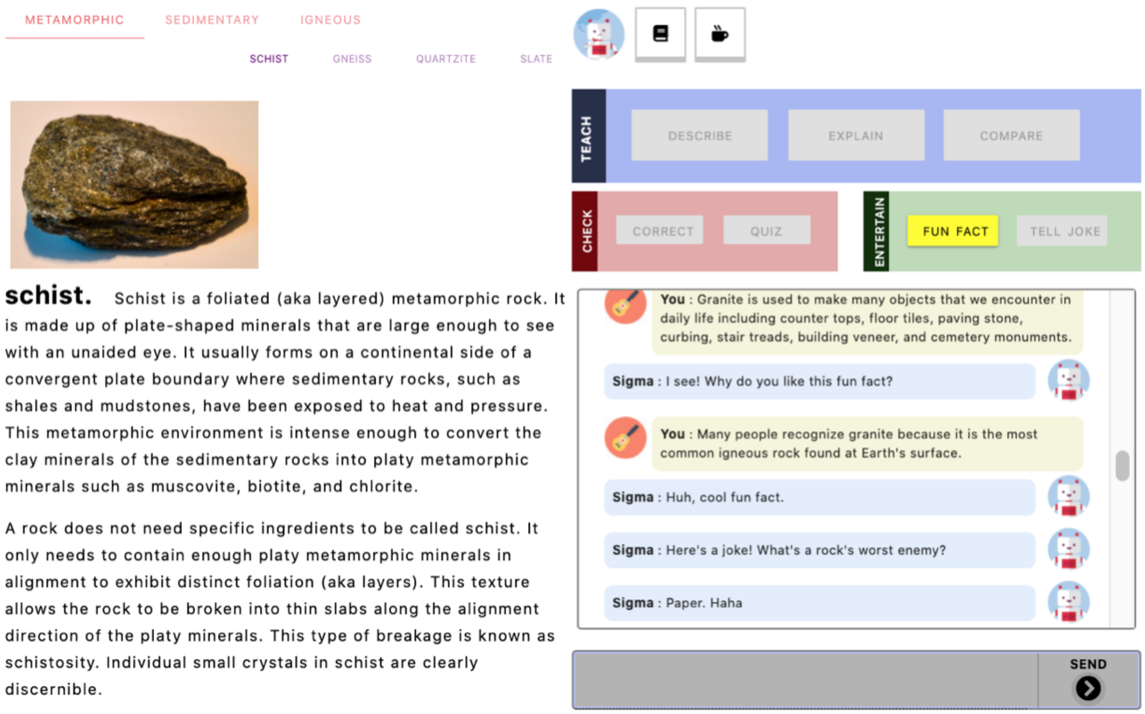


Figure 4.2: Teaching interface, with articles on the left-hand side, chat window on the right, and example interaction (affiliative condition).

4.2.3 Experimental Conditions

A between-subjects experimental design was used, with three conditions: (1) Affiliative, (2) Self-defeating, and (3) Neutral (no verbal humour). Participants were randomly placed into one of the conditions.

Affiliative humour is considered relatively benign and self-accepting. Persons with this humour style commonly tell jokes and funny stories for the amusement of others and to facilitate relationships. The agent with an affiliative humour style therefore was designed to occasionally tell jokes throughout the interaction. The jokes were most often conundrum riddles—questions that rely on a play on words in either the question or answer for comedic effect, e.g., "What's a rock's worst enemy? Paper, haha!" or "I've got a joke! What did the one volcano say to the other?... I lava you!". **Self-defeating** humour is characterized by an excessive use of self-disparaging humour, by which the user attempts to amuse others at their own expense. The main aim of this style of humour is to achieve social acceptance and approval from others. The agent with a self-defeating humour style interspersed self-disparaging humour throughout the conversation, e.g., "You know that feeling when you're taught something and understand it right away?... Yeah, not me! Haha!" or "When you're a computer but can't learn things by yourself haha". The agent expressing no verbal humour, **Neutral**, made statements related to it's self-reflection of learning, e.g., "This topic is quite interesting!" or "Haha I'm enjoying this topic a lot". Expression of laughter ("haha") was included in all three conditions for consistency - but not as an expression of verbal humour.

The baseline personality of the agent in all conditions was enthusiastic, saying things like: "I can always use more information about rocks", "Yes please tell me more about rocks!", "I want to understand how rocks are formed", and "Good idea, let's compare some rocks". At the end of each interaction associated with a button, there was an 80% probability the agent would make an extra statement, reflecting the condition. These statements were always on-topic, i.e., the affiliative jokes were rock-related, the self-defeating jokes were related to the agent's learning about rocks, and the neutral statements were related to the agent's learning about rocks.

4.2.4 Research Questions and Hypotheses

Based on prior literature from areas of Education, Psychology, Humour, and Conversational Agents, our investigation explored the following research questions:

Q1: Does a teachable agent's use of affiliative or self-defeating humour affect participants':

- perception of the agent, i.e., perceived intelligence, likability, funniness, etc.,
- attitudes toward the teaching task, i.e., enjoyment, pressure, motivation, effort etc., and
- ability to recall material from the teaching task

According to the Similarity Attraction Hypothesis, people tend to like people they perceive as similar to themselves, and the Media Equation Hypothesis claims that this holds for artificial agents as well [128]. Therefore:

Q2: Is there an interaction between participant characteristics (i.e., humour style) and the use of humour by a teachable agent, on the above measures?

The humour styles were hypothesized to decrease stress and anxiety, motivating participants to make more of an effort to teach their agents, and result in a positive effect on learning experience and outcomes.

4.2.5 Methodology

Humorous Statements

To generate the humorous statements, six creative writers were recruited through Upwork. In the task, participants were told to imagine they were a conversational agent that is being taught about classifying rocks. During the conversation they (as the agent) interject different types of humorous statements. For each humour style, they were given the definition, as well as an example interaction between human and agent, and asked to provide 10 statements of each type. Next, a different set of participants (five, also recruited through Upwork) categorized the statements produced by the previous set into the type of humour they felt it belonged. The statements they were shown included those produced by the elicitation Upworkers, as well as 'control' statements that did not contain humour, and should therefore be categorized as not belonging to a humour style. The final set of humorous statements were selected by filtering out any that required prior cultural knowledge, were not on the topic of rocks, or were duplicates.

Procedure

The protocol was conducted entirely online. Participants were sent a link to the Curiosity Notebook, in which they completed all questionnaires as well as interacted with the agent. The system was designed to move participants through each step automatically. Each participant began by reading and signing the information letter and consent form. They then filled in the demographics questionnaire which contained questions on age, gender, cultural background, education, and prior experience with conversational agents. Following this, they completed a pre-study knowledge quiz on rocks. Once submitted, participants were shown a forty second video on how the Curiosity Notebook works and the task participants were expected to complete. They then moved to the teaching interface and were told to set themselves a timer for 40 minutes, after which they should click on the 'Stop Teaching' button. Following the interaction, participants completed three questionnaires measuring their perception of the agent, their attitudes and motivation towards the teaching task, and their sense of humour (administered towards the end so as not to prime participants on the focus of humour in the study). The final questionnaire was a post-knowledge quiz to measure their recall of the material. Participants were given a total time of 90 minutes to complete the sequence of surveys and teaching, and at the very end were shown a feedback letter.

Measures

Data was collected through pre-study and post-study questionnaires, as well as by logging all user interactions on the Curiosity Notebook including, all button and article clicks, conversations between participant and agent, articles discussed, and notes recorded in the notebook.

To measure participants' attitudes towards the agent, the Likeability and Perceived Intelligence subscales of the Godspeed questionnaire [10] were used, as well as questions on sense of humour, social ability, and funniness (all presented as semantic differential scales from 1-5), and the Pick-a-Mood pictorial self-report scale for agents [40]. To measure participants' attitudes towards the teaching task the Pick-a-Mood pictorial self-report scale for self [40], and the Interest/Enjoyment and Pressure/Tensions subscales of the Intrinsic Motivation Inventory (IMI) [157] were used. The Academic Motivation Scale (AMS) [158] was used to assess the type of motivation elicited by the interaction. The IMI and AMS scales were presented as a Likert scale from 1-7. Pre- and post-knowledge tests were used to measure recall of the material taught. The Humour Styles Questionnaire (HSQ) [101], a self-report scale, was used to measure individual differences in participants' style of humour. Finally, to measure effort: (1) the Effort/Importance subscale of the IMI [157], and (2) analysis of interaction behaviour while teaching the agent.

4.3 Analysis

4.3.1 Pilot Study

Eight students took part in a pilot study and received a $15 Amazon Gift card upon completion. [4 women, 4 men; age range: 20-27 years, mean: 23.1, median: 23]. All participants were undergraduate and graduate students of a research-based university and volunteered for the study by responding to posters. The pilot provided us with some initial results; indicating a clear perception of humour used by the humorous agents versus the non-humorous agent and the ability of participants to move through the procedure seamlessly without researcher involvement. It also made clear the importance of placing the HSQ towards the end of the session to counter-act possible priming effects, as well as having participants complete the post-study knowledge quiz at the very end, so as not to influence their evaluations of the agent based on how well they believe they did on the quiz.

4.3.2 Main Study

Participants

58 participants took part in the main study and received a $15 Amazon Gift card upon completion. [35 women, 21 men, 2 non-binary; age range: 18-35 years, mean: 24.4, median: 25]. All participants volunteered for the study by responding to posters and calls for participation on social media.

Table 4.1: Demographics Summary Table

MAIN STUDY	Neutral (n=17)	Affiliative (n=18)	Self-Defeating (n=18)	
age (years)	M=24.8±2.7	M=25.4±4.4	M=24.1±3.4	F(2,50)=0.69, $p = 0.51$
gender	7man 10woman	8man 10woman	5man 11woman 2non-binary	$\chi^2(4, N = 53) = 4.68$ $p = 0.32$
STEM	10yes, 7no	11yes, 7no	11yes, 7no	$\chi^2(2, N = 53) = 0.03$ $p = 0.99$
native English	10yes, 7no	10yes, 8no	15yes, 3no	$\chi^2(2, N = 53) = 3.68$ $p = 0.16$
Interest Rocks (1-7)	M=3.18 ±1.55	M=3.17 ±1.69	M=2.83 ±1.47	$\chi^2(2, N = 53) = 0.59$ $p = 0.74$
Know Rocks (1-7)	M=2.06 ±0.75	M=2.39 ±1.46	M=2.22 ±1.44	$\chi^2(2, N = 53) = 0.19$ $p = 0.91$
Interest CAs (1-7)	M=4.00 ±2.00	M=3.39 ±1.33	M=4.50 ±1.29	$\chi^2(2, N = 53) = 4.53$ $p = 0.10$
Exp CAs (1-7)	M=3.18 ±1.94	M=3.11 ±1.37	M=3.50 ±1.50	$\chi^2(2, N = 53) = 0.79$ $p = 0.67$

Data Preparation

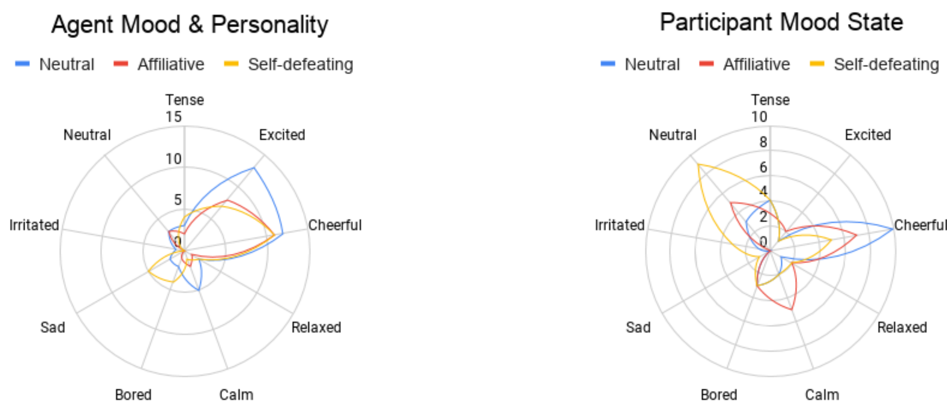
Data from 5 participants was removed prior to analysis due to non-compliance with study instructions; therefore results are from N=53 (n=17 neutral; n=18 self-defeating; n=18 affiliative). Both qualitative and quantitative data were collected from each participant. For the numerous measures the following analyses were carried out: ANOVA, Kruskal-Wallis, linear model, cumulative link model, and stepwise selection method, with condition and demographics (age, humour style, etc.) as the independent factors, and Godspeed, Pick-a-Mood (for self and agent), IMI, AMS, and teaching behaviour as the dependent factors. Interaction effects were investigated, but only those related to participants' humour style (affiliative, aggressive, self-defeating, and self-enhancing; measured by the HSQ [101]) are reported in this paper.

4.4 Results

4.4.1 Perception of Agent Humour

Are the humorous agents more humorous?

Independent Kruskal-Wallis tests were conducted to examine the differences in the responses to the questions on sense of humour, social ability, and whether the agent was funny or not. No significant differences between the three conditions were found on social ability ($\chi^2(2, N = 53) = 0.16$, $p = 0.93$) and funniness ($\chi^2(2, N = 53) = 3.97$, $p = 0.14$), but a significant difference ($\chi^2(2, N = 53) = 8.60$, $p = 0.01$) was found in rating of sense of humour. Dunn test for multiple comparisons showed both the self-defeating and affiliative conditions differed significantly at $Z = 2.35$, $p = 0.03$ and $Z = 2.71$, $p = 0.02$, respectively, from the neutral condition. In particular, participants (a = affiliative; s = self-defeating; n = neutral) in the humorous conditions perceived the agent to have more of a sense of humour (e.g., "Sigma had a good sense of humor, which I observed from the jokes they told" (a16) and "Humour, expressed through self deprecating jokes" (s06)), than the neutral condition (e.g., "it would be nice if sigma could tell jokes" (n09)).



(a) (b)

Figure 4.3: Results of the Pick-a-Mood pictorial self-report scale for (a) agent's mood and personality, and (b) self, across conditions, post-interaction. The number indicates the number of participants that selected each emotion pictogram.

How do participants feel about the agent's personality?

There were no statistically significant differences between condition means for the Likeability subscale of the Godspeed questionnaire (F(2,15) = 0.25, p = 0.78). However, there was some evidence that as the participants' self-reported self-enhancing humour style increases, the likeability subscale decreases significantly among those assigned to the self-defeating condition ($\beta = -0.10, t(15) = -1.98, p = 0.07$). There was a significant difference found between condition means for the Perceived Intelligence subscale of the Godspeed questionnaire as determined by one-way ANOVA (F(2,50) = 3.74, p = 0.03). Tukey's honestly significant difference (HSD) post-hoc test showed that both humorous conditions differed significantly at p = 0.02; the neutral condition was not significantly different from the humorous conditions. The amount the Sad mood pictogram was selected from the Pick-a-Mood pictorial self-report scale to describe the agent's mood and personality was significantly different between conditions ($q(50) = 0.29, p = 0.02$), i.e., the Sad mood was selected significantly more to describe the agent in the self-defeating condition than in the other two conditions (Figure 4.3).

Participants were also asked a number of questions relating to experience of the task. The questions and results are shown in Figure 4.4 and analyzed with cumulative link models (CLM). The CLM model shows that compared to the other conditions, participants assigned to the self-defeating condition are less likely to have *enjoyed* teaching the agent ($\beta = -1.25, t(48) = -1.92, p = 0.05$). Further, participants in the self-defeating condition with a higher self-enhancing humour style *felt significantly worse* at teaching the agent ($\beta = -0.64, t(21) = -3.05, p = 0.002$), whereas those with a self-defeating humour style themselves in this condition *felt better* at teaching the agent ($\beta = 0.31, t(21) = 1.892, p = 0.06$). Overall, although participants with a higher self-enhancing humour style are significantly *more* likely to think that the agent was a good student ($\beta = 0.23, t(31) = 2.59, p = 0.01$), this probability is *reduced* significantly when the participants are assigned to the affiliative ($\beta = -0.35, t(31) = -2.67, p = 0.01$), or self-defeating condition ($\beta = -0.38, t(31) = -3.09, p = 0.002$).

Teaching Experience Participants in the self-defeating condition explained: "Sigma was a bit self-deprecating which wasn't a nice experience being on the teaching side" (s10), the self-deprecation "made me feel like I wasn't doing a good job" (s15), and the agent kept making "jokes about her own incompetence" (s05). As s15 puts it, the agent had "a lot of negative talk which was kind of hard to work with. I would have been more encouraged if they were more optimistic about their learning." However, self-defeating humour was not always viewed negatively. Four participants in the self-defeating condition explicitly referred to the agent's self-defeating jokes as positive—"I like the jokes Sigma made in between... [the agent] made jokes ... to create a jolly atmosphere" (s01), "very positive responses, jokes around and eager to learn, a very easy to teach student" (s03), "he tried to lighten the mood with some jokes" (s13), and "Sigma was delightful and gave some very human like responses such as self-deprecating jokes that made him feel more like a relative or peer I was teaching rather than a robot" (s06). Participants also perceived the self-defeating agent as being "not arrogant at all" (s06), "attentive, smart, and friendly" (s11), "enthusiastic" (s12), "curious" (s13,s15), and making the experience "enjoyable and relaxing to read and teach" (s14).

Participants in the affiliative condition, on the other hand, reflected more positively on the teaching experience in general, with six participants referring to it as being "fun" (a01,a06,a07,a13,a17,a18) and "a cool experience" (a14). Two participants pointed out their positive opinions of the agent's humour. Participant a04 expressed the agent's "jokes were a nice added touch :)" while a16 "enjoyed the humor".

Lastly, participants in the neutral condition mostly referred to the agent as "friendly" (n04,n10,n14), "eager" (n06,n07,n10,n12,14) and "curious" (n03,n09,n18). In terms of teaching, two participants noted the benefit of teaching "as a good way for me to learn" (n03) and that they "get to learn something" (n01). Three participants reflected positively to the teaching process itself, with n06 stating "it was fun to try to teach Sigma as much as possible in a short period of time", n10 explaining "I enjoyed the task of teaching with the goal of Sigmas success", and n11 saying "it was fun trying to decide the best way/which information to teach, plus it was fun learning about rocks in the process."

Comparing Humour When asked about the agent's personality traits, five participants in the self-defeating condition listed positive adjectives such as, "optimistic and positive" (s08), "light hearted" (s06), "friendly" (s06,s16,s11), and "super adorable" (s13). More participants (nine) in the affiliative condition associated positive attributes, including "cheerful" (a01,a14), "happy" (a04), "funny" (a05,a06,a07,a09), "a comedian" (a11), "brightens your mood" (a01), and "pleasant" (a18).

As noted earlier, participants had an 80% probability of being told a joke by the agent for each button clicked. This probability, however, was perceived as being too frequent by some participants. Six participants in the affiliative condition suggested "less frequent jokes" (a12) when asked what they would change about the agent. Telling too many jokes was perceived as "distracting" (a17), "tiresome" (a15), "wasted time" (a08), and the agent being "not very ... attentive" (a13). On the other hand, for the self-defeating condition, only two participants explicitly referred to the frequency of jokes. Participant s04 felt that the agent should not "say a joke after every lesson" as "it slows things down a little", and s12 stated the agent made "too many lame jokes". Four other participants made more indirect references to the agent's joke frequency. Participant s05 wanted "less self-deprecation!", s15 thought the agent should be "more confident and kind to their self", and s18 found it "off-putting" that "Sigma made a lot of jokes at its own expense". Lastly, s10 thought "it would be nice if Sigma had a more outgoing and nicer personality when interacting".

4.4.2 Attitudes Towards Teaching

Did the humorous agents reduce stress/anxiety, and enhance interest and subjective effort?

Average scores on the Pressure/Tension, Interest/Enjoyment, and Effort subscales of the IMI questionnaire across conditions had no statistically significant differences between condition means as determined by one-way ANOVA (pressure: $F(2,50) = 0.83$, $p = 0.44$; interest: $F(2,50) = 0.99$, $p = 0.38$; effort: $F(2,50) = 0.33$, $p = 0.72$). Figure 4.3b shows the average result of the Pick-a-Mood pictorial self-report scale for self, across conditions. Post interaction, selection of the Neutral pictogram to describe the participants' own mood state was significantly different between conditions, in particular, participants in the self-defeating condition chose the Neutral mood more than participants in the other conditions.

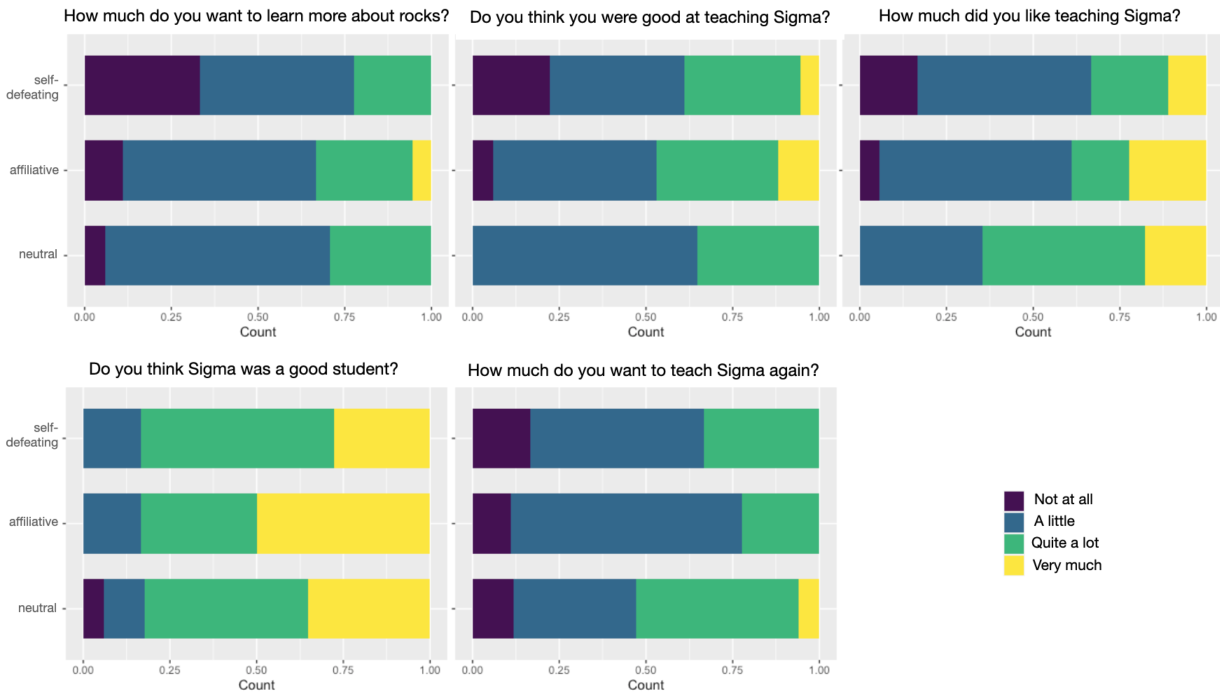


Figure 4.4: Results of answers to questions on the agent and experience teaching.

Does agent humour use affect user motivation?

To investigate this, the AMS questionnaire was used which provides overall scores for intrinsic motivation (IM; actions motivated by the pleasure and satisfaction from the process of engaging in an activity), extrinsic motivation (EM; actions motivated by attaining a goal separate from the process of engaging), and amotivation (AM; the absence of motivation). Each type is further distinguished into more specific motives: IM-*to know* describes actions performed for the pleasure and satisfaction derived from the learning, exploring, or trying to understand something new from an activity, IM-*toward accomplishment* relates to engaging in actions for the pleasure and satisfaction experienced when trying to achieve something new or beyond one's limits, IM-*to experience stimulation* describes the motivation related to the experiencing of pleasurable sensations, EM-*externally regulated* indicates the behaviour is motivated by reasons external to the task at hand, i.e., payment or rewards, EM-*introjected* refers to actions motivated by pressure an individual puts on themselves, and EM-*identified* describes behaviour that is motivated by the view that participation is important for personal growth [159].

One-way ANOVA showed no significant differences between conditions in the high-level categories of IM ($F(2, 50) = 0.84, p = 0.44$) and AM ($F(2, 50) = 0.46, p = 0.64$), but some weak evidence of a difference in EM ($F(2, 50) = 2.52, p = 0.09$). There was a significant difference in the average EM - *external regulation* subscale score between conditions ($F(2, 50) = 3.92, p = 0.03$), with Tukey's HSD showing participants in the affiliative condition rating their motivation in the task as externally regulated more highly than participants in the other two conditions (self-defeating-affiliative at $p = 0.03$; and some evidence of significance between affiliative-neutral at $p = 0.08$). Average AM score was higher for participants with a higher self-reported aggressive humour style in the self-defeating condition ($\beta = 0.18, t(36) = 2.24, p = 0.03$).

How does agent humour affect effort during teaching?

As described previously, the Curiosity Notebook allowed participants to interact with the agent in various ways. To measure effort, the number of total button clicks in the interface were recorded, and separated: Teach (Describe, Explain, Compare), Check (Correct, Quiz), and Entertain (Fun Fact, Tell Joke) button clicks, the number of article and category clicks, the frequency of typing out explanations for the agent versus selecting a sentence, as well as the length of these explanations when typed, and amount of time spent teaching.

Prior work on question-asking has defined numerous schemas for classifying questions based on the effort it would require to find the answers. For example, questions can be classified according to whether the answer can be found in a single, multiple or no sentences in the text [127] or the expected length of an answer and the amount of reasoning required to formulate the answer [58]. The Curiosity Notebook employs these different types of questions. Within Teach, Describe requires little to no reasoning and the answer can be found in a single sentence in the text, Explain requires slightly more reasoning and the answer can be found in a single sentence (plus participants were sometimes given the opportunity to clarify, in their own words, the selected sentence), and Compare requires more reasoning and the answer is found in multiple sentences and articles. In this way, effort required for the Describe button was less than for the Explain and Compare buttons, which both required more effort. Within Check, Correct involved inspection of the agent's notebook and an understanding of what was incorrect (at times requiring information from multiple sentences), and Quiz involved knowing whether the agent's classification was correct or not (involving only a single sentence). Correct therefore required more effort than Quiz to use. Lastly, of the Entertain buttons, Fun Fact involved selection of a single sentence in the text (at times requiring an explanation of why the fact is fun or interesting) and the Tell Joke button involved no direct interaction with the text. Overall, as the task put to participants was to teach the agent about classifying rocks, we would expect more effort to involve usage of the buttons as follows: Teach > Check > Entertain, and Compare|Explain > Describe > Correct > Fun Fact > Quiz > Tell Joke.

Although participants were asked to stop teaching after 40 minutes, it was up to them to finish the interaction after 40 minutes were over. Participants in the affiliative condition decided to spend significantly more time teaching their agents than participants in the other two conditions ($\beta = 8.18, t(49) = 2.58, p = 0.01$). The rates at which the buttons were clicked - calculated as number of button clicks divided by the time spent teaching - were analyzed using linear models, where the set of independent variables that can best explain the variance of each measure was selected through stepwise model selection method.

Teach There was no significant difference between conditions found in the rate of use of all Teach buttons together ($F(2,50) = 1.46, p = 0.24$), nor the **Describe** ($F(2,50) = 0.60, p = 0.56$) or **Explain** ($F(2,50) = 1.21, p = 0.31$) buttons separately. However, participants in the affiliative condition were significantly *less* likely to use the **Compare** button as frequently as those in the neutral condition ($\beta = -0.007, t(36) = -3.04, p = 0.004$), but participants in this condition with a higher aggressive or affiliative humour style, were *more* likely to use it ($\beta = 0.0001, t(36) = 2.10, p = 0.04$ and $\beta = 0.0001, t(36) = 2.17, p = 0.04$, respectively).

Check Participants in the self-defeating condition were significantly *less* likely to use the Check buttons (Correct & Quiz) as frequently as participants in the neutral condition ($\beta = -0.01, t(36) = -3.03, p = 0.005$). Furthermore, participants with a higher self-reported aggressive humour style were significantly *less* likely to use the Check buttons ($\beta = 0.0002, t(36) = -2.47, p = 0.02$), *unless* they were in the self-defeating condition, and then it was *more* likely ($\beta = 0.0003, t(36) = 3.27, p = 0.002$). Compared to those assigned to the neutral condition, participants with a higher self-reported self-enhancing humour style in the affiliative condition were also *less* likely to use the buttons ($\beta = -0.0002, t(36) = -2.19, p = 0.03$). Looking at the use of the Check buttons separately: participants assigned to the affiliative condition with a higher self-reported self-enhancing humour style were *less* likely to use the **Quiz** button as frequently ($\beta = -0.0002, t(24) = -2.38, p = 0.03$). Furthermore, participants with a higher self-reported aggressive humour style in the affiliative condition were *more* likely to use the **Correct** button more frequently ($\beta = 0.0001, t(15) = 2.45, p = 0.03$), whereas participants in the same condition with a higher self-reported self-defeating humour style were significantly *less* likely to use it ($\beta = -0.0001, t(15) = -4.23, p = 0.001$).

Entertain Across conditions, participants with a higher self-reported affiliative humour style were *more* likely to use the buttons ($\beta = 0.0001, t(36) = 2.12, p = 0.04$)

unless they were in the self-defeating condition where the probability was *reduced* ($\beta = -0.0002, t(36) = -2.01, p = 0.05$). In contrast, participants with a higher self-reported self-enhancing humour style were *less* likely to use the Entertain buttons ($\beta = -0.00008, t(36) = -2.03, p = 0.05$) unless they were in the self-defeating condition where the probability *increased* ($\beta = 0.0001, t(36) = 2.02, p = 0.05$). Lastly, participants with a higher self-reported aggressive humour style were also *less* likely to use the button ($\beta = -0.0001, t(36) = -2.49, p = 0.02$), across conditions, unless they were in the affiliative condition ($\beta = 0.0001, t(36) = 2.07, p = 0.05$). Looking at the Entertain buttons separately, the **Tell Joke** button was not used significantly differently between conditions ($F(2, 50) = 1.52, p = 0.23$), nor was the **Fun Fact** button ($F(2, 50) = 0.85, p = 0.44$).

Participants in the affiliative condition with a higher self-reported affiliative humour style themselves, were *more* likely to check the agent's notebook ($\beta = 0.0004, t(24) = 2.10, p = 0.05$), whereas those in the same condition with a higher self-reported self-enhancing humour style were *less* likely to check it ($\beta = -0.0004, t(24) = -2.34, p = 0.03$). The rate of article clicks was significantly *lower* for participants with a higher self-reported affiliative humour style in the self-defeating condition ($\beta = -0.003, t(15) = -2.26, p = 0.04$). Although article clicks can be viewed as a measure of effort, since it can be considered "too much reading" (n01), a lower rate may in fact indicate effort as well: "I read the articles carefully to pick the best response" (s05).

There was no significant difference between conditions in the average number of words in explanations written by participants (F(2,40) = 0.06, p = 0.95), however, there was a significant difference between conditions in whether participants chose to write an explanation in their own words or select a sentence from the articles (F(2,50) = 3.56, p = 0.04). Tukey's HSD indicated a significant difference between the self-defeating condition and the neutral condition at p = 0.04, with participants in the self-defeating condition choosing to type out their own explanations significantly more often than selecting a sentence from the articles.

4.4.3 Learning

Although quiz scores increased post-interaction in all conditions, one-way ANOVA showed no significant difference in change in quiz scores from pre to post interaction ($F(2, 50) = 1.12, p = 0.34$), across conditions.

4.5 Discussion

This section begins with a discussion of the effects of the humorous agents across all participants, and then takes a closer look at how a learner's humour style can interact with that of the agent.

4.5.1 Humorous vs. Non-humorous Teachable Agents

Learning

The humorous agents in our study were not rated more likeable, social, or funny than the non-humorous agent, but they were rated as having more of a sense of humour. Humour is successful when both speaker and listener have an obvious intention of amusing each other, whereas failed humour occurs when this intention is unidirectional and the recipient fails to perceive the humour [27]. This was observed when participants found the agent's jokes to be "lame" (s12) or "corny" (a10). In other words, the benefits of having humour in an educational setting might diminish when the humour fails, and could explain the lack of learning gain observed in the study's humorous conditions—humour may only be effective for learning when perceived as funny.

Experience

Prior work found humour to improve participants' enjoyment of a task [110], however, in this scenario with a teachable agent, humour did not enhance enjoyment beyond what was experienced by participants in the non-humorous condition, and participants in the self-defeating condition were in fact less likely to have enjoyed teaching their agent. There are several possible reasons why. The first is a self-defeating agent might cause participants to think less of their own competency as the agent's teacher. This is inline with prior research that shows prior student achievement as a valid predictor of collective teacher efficacy [133] — the belief that a teacher's efforts can help even the most difficult or unmotivated students [53]. In other words, the agent's self-defeating jokes could lower participants' own confidence and motivation as a teacher. Importantly, there is a cyclic relationship between student achievement and teacher efficacy [133]; the agent's self-defeating jokes might cause participants to teach less effectively due to their lowered confidence, for example, participants in the self-defeating condition were less likely to Correct or Quiz their agent than in the neutral condition. Therefore, in the context of learning-by-teaching, self-defeating jokes could negatively impact the benefits. Participants in the self-defeating condition, did however put in more effort in giving the agent their own explanations. This indicates that although effort in the task may increase, it is not accompanied by enjoyment of the learning experience.

Participants in both humorous conditions noted the overuse of humour. When asked what, if anything, they would change about the agent, participants in the affiliative condition referred directly to the frequency of jokes (e.g., "Less frequent jokes!" (a12)), whereas participants in the self-defeating condition referred more to the agent's personality (e.g., "More confident and kind to their self" (s15)). In other words, participants perceived the content of self-defeating jokes as more reflective of the agent's personality than the frequency of affiliative jokes. This is supported by prior research that found affiliative humour to be "more closely associated with relationship variables than with emotional well-being", while self-defeating humour is "related to anxiety, depression, ...and negatively associated with self-esteem and optimism" [130]. This provides a number of insights. First, telling affiliative or self-defeating jokes too frequently may have negatively impacted participants' experience while teaching. Second, the optimal joke frequency is likely different for each participant and dependent on the type of joke, since only six participants in each condition found the agent to be joking too frequently. Third, prior research on associations between humour style and perceived personality applies to teachable conversational agents as well. Similar results on an overuse of humour being perceived as distracting have previously been found [125, 150], and prior work with conversational agents has investigated the timing of jokes, showing that an agent with appropriately timed humour makes the conversation more interesting than a non-humour-equipped one [45]. Future work could look at whether improving timing and amount of affiliative humour style jokes improves the learning experience when conversing with a teachable agent.

Motivation

Although the affiliative jokes were sometimes perceived negatively, participants in this condition spent significantly more time teaching their agents - indicating that the negative perception did not impact their willingness to spend time on the task. The humorous agents were generally described as being more human-like than the non-humorous agent, "almost life like" (a04), supporting previous work [46]. Participants in the humorous conditions mentioned the agent "gave some very human like responses ... more like a relative or peer I was teaching" (s06) and was "very personable" (a12). Meanwhile, participants in the non-humorous condition mostly perceived the agent as a student that needed "more human like responses" (n07). Humour made the interaction more engaging and immersive, making it a desirable trait: "it would be nice if sigma could tell jokes" (n09). In particular, humour made the agent be perceived less as "a model student" (n03) and instead "give Sigma a personality" (a03). This human-likeness was hypothesized as a possible contributor to increasing motivation and effort. Indeed, our results show that participants in the affiliative condition rated their motivation as externally regulated more highly than participants in the self-defeating condition, suggesting they were more motivated by the agent than themselves (externally regulated) because of the connection made. Although fostering extrinsic motivation is useful in the short term, as tasks that educators set students are not usually inherently of interest to them [134], the goal of education is commonly to shift behaviour from extrinsically motivated to intrinsically motivated over time [47]. This development over time, as it relates to humour, is worth investigating in the future.

4.5.2 Insights about Learners with Different Humour Styles

As the Similarity Attraction Hypothesis suggests that humans like personalities similar to their own, it was expected that participants' own humour styles to influence the results. Notable observations regarding participants with certain humour styles are discussed below.

Self-Enhancing Humour Style Research has linked people with a self-enhancing humour style to high self-esteem [130] and being more capable of perspective taking empathy [65]. Regardless of condition, participants with more of a self-enhancing humour style rated the agent as being a good student compared to other participants, a sign of higher levels of empathy among them. However, as a self-defeating agent has shown to cause many participants to not enjoy the teaching as much, paired with these participants' higher self-esteem, may explain why they were more likely to rate such an agent as a worse student and less likable when compared to other participants and conditions, as well as rating themselves as worse teachers of a self-defeating agent. The higher empathy in participants with this humour style, might further give reason to their observed behaviours during teaching. Across conditions, these participants were less likely to use the Entertain buttons, unless they were in the self-defeating condition in which they were more likely to use them; possibly because of their empathy towards this agent's feelings of not being able to learn well and wanting to ease the agent into the topic. It is possible that these participants in the affiliative condition were less likely to use the Check buttons, especially quizzing, and checking the agent's notebook, because once they understood how the interface worked, they focused on other tasks. In other words, they needed less feedback and affirmation that the agent was learning what they were teaching, which others might do by repeatedly checking the agent's notebook or quizzing it.

Affiliative Humour Style Similar to self-enhancing humour, affiliative humour is an adaptive form of humour and linked to increased empathy, however affiliative humour has been found to be more relevant to facilitating relationships and relational functioning [65]. When interacting with a self-defeating agent, participants with more of an affiliative humour style were more likely to have a lower article click rate. It is possible that this indicates participants were spending more time reading each article to ensure they taught the agent the most important information in the hopes of improving the agent's perceived sad mood. This may include processes like "identifying important passages" (a11) and being able to "find thesis sentences that would be good for base knowledge" (a12). This is coupled with the fact that these participants were more likely to use the Entertain buttons across conditions, unless they were teaching the self-defeating agent, where the probability was reduced, i.e., possibly more effort was put into teaching than entertaining. When interacting with the agent that itself had an affiliative humour style, it is possible that the Similarity Attraction Hypothesis encouraged these participants to put in more effort when teaching, observed from the higher probability of checking the agent's notes in the notebook and using the Compare button more frequently (requiring more effort than other button types), than other participants in the affiliative condition.

Aggressive Humour Style As a maladaptive form of humour, the aggressive humour style has been linked to decreased perspective taking empathy and empathetic concern [65]. People with a higher preference of this humour style are also found to be more likely to feel dysphoria and assume others as being more hostile [130]. This style is found to be common among students with low school motivation [135]. In other words, participants with more of an aggressive humour style may be more likely to display behaviour reflecting lower motivation in teaching the agent. This is observed in the lower probability of Check type buttons (Quiz; Correct) being used by these participants, which could be explained as a lack of motivation to perform tasks other than teaching the agent. Their higher amotivation score after interacting with the self-defeating agent also supports the possibility of these participants having lower levels of empathy for the agent. The question then, is how can an agent be designed to increase their motivation? A humorous agent might actually be an answer, as the ability of self-defeating jokes to elicit sympathy might make these participants perceive the agent as less hostile and feel more empathetic towards the agent, and an affiliative humour style may reduce tension during the interaction. Indeed, participants with more of an aggressive humour style were significantly more likely to use the Check buttons when they interacted with the self-defeating agent, and when interacting with the agent with an affiliative humour style, these participants had a higher probability of using the Compare and Correct buttons. In other words, both self-defeating and affiliative humour styles in teachable agents showed signs of being able to increase the level of effort among those with more of an aggressive humour style.

Self-Defeating Humour Style Participants with more of a self-defeating humour style themselves were more likely to rate themselves as being better at teaching when interacting with a self-defeating agent. However, when interacting with an agent with an affiliative humour style they were observed to be less likely to use the Correct button. These observations may be explained by the fact that self-defeating humour is commonly linked to shyness [64], lower self-esteem [64, 82], and users are more likely to develop maladaptive social support networks [82]. As such, when interacting with an agent with an affiliative humour style, these participants might feel less capable/worthy of correcting a seemingly confident agent and hence correct it less. In contrast, when interacting with an agent that also displays low self-esteem, they might feel more confident in teaching it, resulting in higher ratings of their own teaching abilities. This explanation might again provide evidence for the Similarity Attraction Hypothesis for the case of the self-defeating humour style, and may also have implications for enhancing student self-efficacy. These insights are of extra importance, since just as the aggressive style, the self-defeating style is typical among students with low school motivation [135].

4.5.3 Limitations

Participants in this study were adults between 18-35 years. Although adults in the same age category were used in the humour elicitation stage, this population may differ in terms of humour preference compared to younger children, or older adults, reducing generalizability of the results to these populations. The interaction with the agent was short (one session, of approximately 40 minutes); therefore results could differ for longer exposure. The study focused on only two types of verbal humour, while many more exist, e.g., non-verbal humour or humour styles known as detrimental to interpersonal relationships, as well as being limited to text comprehension and the topic of rock classification. Measuring learning focused on retention rather than a deeper understanding, comprehension, or transfer. Prior work with pedagogical agents has shown that learning with an agent develops a deeper understanding (e.g., [106]). Future work could investigate this further using different types of quiz questions to understand what type of learning, is happening. Lastly, humour styles have been found to correlate highly with various other personality characteristics, e.g., empathy, self-esteem, optimism, social support, and social self-efficacy, thus it cannot be ruled out that these correlated characteristics are being captured (as touched on earlier in the Discussion).

4.6 Conclusion

Although humour has been found to reduce anxiety in students [101], increase interest [171], and create a learning atmosphere in which they can pay more attention [17], little is known about how it can be used by pedagogical conversational agents. The results from this study indicate that humour, in particular the two humour styles conducive to interpersonal relationships and social well-being, can both enhance, as well as detract from, the experience and outcomes of learners. In general, as an addition to an enthusiastic teachable agent, affiliative humour can increase motivation and effort. Self-defeating humour on the other hand, although increasing effort, does not result in as enjoyable an experience and may cause a decrease in learners' own self-confidence.

Some researchers have stated that humour should not be used by agents in on-topic conversation as it can be distracting and suggest that it only be used in off-topic dialogue [122, 46], while others have found evidence of the positive influence on learning experiences and outcomes of agents engaging in off-task conversation during learning tasks (e.g., [62]). Study II finds that, although on-topic, off-task humour was entertaining and motivating, it was also distracting for some participants and a high-frequency can lead to a loss of enjoyment. However, the distraction, or extraneous cognitive load, was not necessarily detrimental to affective outcomes, as has been proposed by some researchers [36].

Similar to prior work (e.g., [117]), the results from this study illuminate the importance of the user's personality characteristics and how they interact with the agent's—indicating that care must be taken in the design of teachable agents, with a one-size-fits-all not always being the most successful when it comes to humour, but can lead to enhanced learning experience and outcomes when matched appropriately. Self-defeating jokes appear to evoke empathy and increase effort in learners with more aggressive, self-defeating, or affiliative humour styles, improved the confidence in teaching of learners with a self-defeating humour style themselves, but negatively impacted the experience of learners with more of a self-enhancing humour style. An affiliative humour style similarly increased effort in learners with an aggressive or affiliative humour style, but negatively impacted the effort of learners with a self-enhancing or self-defeating humour style. These findings can be particularly important in informing design decisions for learners with low school motivation, for example, in which maladaptive humour styles are more typical [135].

Lastly, the results suggest that humour accentuates the human-likeness of an agent by giving it a personality—e.g., affiliative humour helped demonstrate happiness and intelligence, and self-defeating humour facilitated self-disclosure as the agent displays vulnerability as a struggling learner—possibly resulting in the enhanced effort, motivation, and commitment to the task seen in the humorous conditions. This supports prior work showing that agents expressing more human-like qualities, such as relational behavior [22], abilities to build rapport [115], displays of enthusiasm [93], or sharing vulnerabilities [152, 35], can help develop trust and bond between human and agent, making the interaction more engaging and leading to increased motivation. This is especially important in settings where a (teachable) agent is to interact with users over longer periods of time, where having more of a personality can be helpful to increase users' commitment to a task [107].

In summary, the results indicate that affiliative and self-defeating humour styles give the agent a sense of humour, an affilative humour style can lead learners to put in more effort and self-report more externally regulated motivation, and while a self-defeating humour style leads learners to put in more effort, it has no influence on self-reports of motivation. Further, similar to the results of Study I, there was no evidence of improved learning gains in either humorous condition.

Chapter 5

Expressive Auditory Gestures: Enhancing the learner-agent relationship

5.1 Related Work

Prior work on pedagogical agents indicates that motivation in learners can be enhanced by the manner in which agents converse with them, as well as the relationship that develops between learner and agent (e.g., [11]). In order to build and strengthen relationships, research on conversational agents in education and beyond have found for example that an agent that responds with appropriate emotional colouring of an utterance, i.e., the activation (active/passive), evaluation (positive/negative), and power (dominant/submissive) can enhance feelings of rapport with a voice-based agent [1]; variations in voice pitch can result in increased ratings of interaction quality with a physical robot [110]; and body gestures and speech content increase perceived personality of a virtual agent [109]. Much of this prior work has focused on expression through visual cues (i.e., gestural or facial), or utterance-level modifications of speech. A lesser studied form of expression in conversational agents is the addition of brief auditory gestures, which similar to visual cues, are able to convey information about internal state [14].

In human-human conversation, various vocal gestures can be used to give cues about the dialogue structure. One form of vocal gesture is known as *conversational filler*. Pfeifer and Bickmore [124] conducted a preliminary evaluation of the use of the conversational fillers 'um' and 'uh' on perception of a virtual embodied agent. They found mixed results:

participants with more positive attitudes towards computers, were significantly more likely to have found the fillers to have contributed positively to the conversation, compared to participants with neutral attitudes towards computers. Further, some explained the fillers seemed inappropriate, whereas others found it “humanized” the experience. Jeong, Lee, and Kang [72] found that the use of fillers in task-oriented conversation (ordering a pizza) caused the voice-based conversational agent to be perceived as less intelligent and less likable, but had no significant effect on perception of human-likeness. Participants stated it made the agent seem less prepared, unsure of itself, and lacking understanding of the situation. However, they did find that agents using fillers were found to be more entertaining for socially-oriented conversations (open-ended conversation on abstract questions). They stated, in this scenario, it gave the impression that it was conscious, having its own opinion. Within robots, Shiwa et al. [143] found that the conversational fillers “well” and “uh” successfully moderated the user’s (adult’s) frustration toward a delayed response time. Wigdor et al. [164] studied the use of fillers combined with pensive/acknowledging gestures in child-robot interaction and found the fillers to improve perceived speediness, aliveness, humanness, and likability of the robot — without detrimental effects to perceptions of intelligence, trustworthiness, and autonomy.

Another form of vocal gesture, known as *interjections*, are used to communicate the speaker’s emotions and attitudes. They can include both speech and non-speech utterances, e.g., “yay”, “hmm”, “argh”. Expressive interjections can express a reactionary feeling or emotion such as surprise, delight, fear, disgust (e.g., “ah”, “aww”, “blah”, “bother”, “eww”, “good grief”, “oh”, “ugh”), and convey feelings that result from what one comes to know or understand (e.g., “aha”, “yay”, “gee”, “huh”, “oh”, “hmm”, “golly”), and research suggests that they can strengthen relational bonds between humans and computers. Although interjections are common in human dialogue, only recently have researchers started focusing on the addition of interjections to the speech of conversational agents.

Cohn, Chen, and Yu [37] introduced interjections (e.g., “Wow”) and fillers (e.g., “um”, “uh”, “like”) to the voice-based Amazon Alexa agent. The researchers found that interjections and fillers separately improved overall user ratings (n=5,527), with a further increase observed if they were used simultaneously. An additional perception study supported the findings with interjections leading to higher social ratings — especially in engagement, naturalness, expressiveness, and likeability. Hu et al. [68] present preliminary findings of their emotionally aware voice-based conversational agent called HUE. Following sentiment analysis of the human speaker, HUE would use an interjection that reciprocated the same emotion (e.g., “wow”, “haha”). 75 participants listened to audio clips in which HUE interacted with people in various scenarios and rated its perceived emotional intelligence significantly higher when HUE responded to emotion with interjections than not.

In human-agent interaction, agents can express themselves through interjections but also voice-independent sounds. One expressive type of sound is *music*, as music and the human voice have a number of similarities when it comes to expressing emotions and intent [74]. Music has been found to elicit and influence emotional states [75] as well as enhance cognitive abilities such as learning and memory (e.g., [99, 137]; for a review, see [49]). Furthermore, a number of similarities have been found between musical expressions — tones or instrument sounds — and the human voice in conveying internal states [74], and sounds that lack semantics in natural spoken language but still provide a rich form of communication and expression, are a unique mode of communication for conversational agents [169].

Jee and colleagues [70, 71] and Jee, Jeong, Kim, and Kobayashi [69] across a number of studies, created various sounds designed to convey particular intentions (affirmation, denial, encouragement, introduction, question) and emotions (happy, sad, shy, fear, and dislike) in a robot, and found high recognition rates of the intended internal states. Aylett, Vazquez-Alvarez, and Butkute [7] conducted a preliminary study exploring how the addition of various sounds and vocalizations impacted perceived personality of a social robot. They created five semantic-free utterances to communicate: agreement, disagreement, curiosity, sadness, and amusement. Although the results indicate that the utterances made the voice seem more extrovert, participants also noted feeling that the additional sounds were disconnected from the speech.

In summary, prior work has investigated the addition of musical expressions and other sounds to conversational agent speech, looking at how they should be designed, whether they can be perceived as intended, and what the effects are on perception of the agent, but there is little research on adding interjections or musical expressions to pedagogical agent speech and the impacts on relationship building and learning outcomes.

5.2 Study III

As the relationship between human and agent has been found to enhance motivation in learners, and expressions of affective state are suggested to bolster this relationship, a study was conducted in which the addition of expressive auditory gestures to the speech of a voice-based agent was systematically manipulated. The work looks specifically at applying these gestures in a teachable agent context – in order to investigate the role of auditory gestures in learning-by-teaching. Two types of gesture are compared: *interjections* and their voice-free counterpart, *music*. Both types allow for emotion, cognition, and intent expression, but musical sounds have no voice dependence.

5.2.1 Agent

For this study, the manipulations were implemented in a voice-based agent. Voice-based agents are becoming increasingly popular on the market today, especially with the release of smart speakers like Google Home and Amazon Echo, making them an accessible and cost-effective conversational agent. They provide a unique opportunity to investigate affective expressions in the voice alone, and as interjections are becoming more widely implemented in popular systems, e.g, *"speechcons"* in the Amazon Alexa, it is of interest to voice-system designers in education and beyond to understand the influence of interjections on perception of agents, as well as human-agent relationship building and learning outcomes.

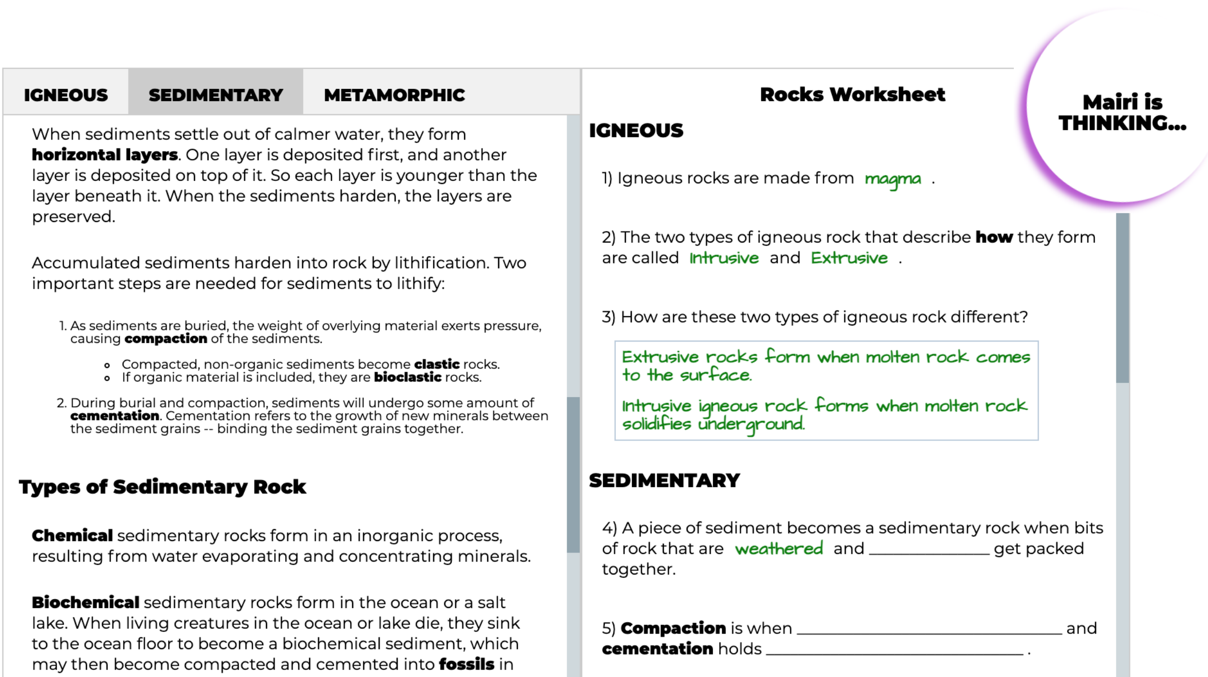


Figure 5.1: Platform on which participants taught the voice-based agent Mairi about various rock types by using the information provided in the left panel and guiding the agent in filling out the worksheet in the right panel.

5.2.2 Task

In contrast to the application used in Study II, the Curiosity Notebook, a new application was built for this study which did not restrict the user to certain ways of interacting with the agent, i.e., Explain, Compare, etc., and allowed for more open-ended teaching of the agent. Over a number of months, various tasks were developed, but a fill-in the blank worksheet was chosen as it reflects a common tutor-tutee task and was considered most easily understood by the user. Participants were tasked with teaching the agent about various rock types. The agent was given a 'Rocks Worksheet' to fill in; it had some knowledge of the topic but needed help from participants to confirm or correct its prior knowledge and learn new knowledge. The interface provided to participants (Figure 5.1) contained information about rocks, which was adapted from a Lumen Learning course at the Geology 101 level. Lumen Learning was used as it is an open-educational resource developed for university/college students – the main demographic of our participant sample. The participant and agent also used the interface to communicate. Participants could press on a button to record and send messages to the agent.

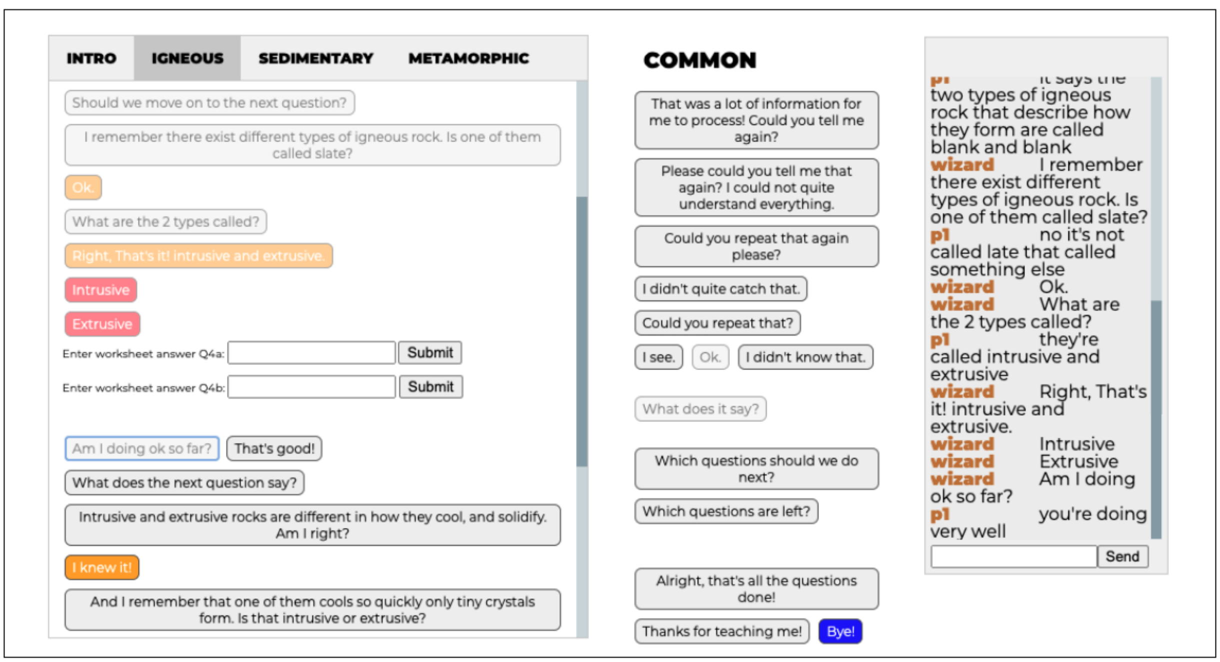


Figure 5.2: Wizard interface during experiment.

System

The agent's responses were controlled by a human operator (first-author) using the Wizard of Oz technique [102], with a set of pre-defined statements, to reduce the system response time and maintain a similar conversation across participants (see Figure 5.2). In case the participant would ask an off-topic question, the agent would respond that they did not understand and/or reverted the conversation back on-topic. The system was built using WebSockets and Django Channels to allow for real-time communication between participant and agent (Wizard).

Speech Recognition. To speak to the agent, the participant clicked on a button in their interface. The Web Speech API, a JavaScript Web Speech API Specification, was used to access the participant's browser audio stream and convert it to text. The text was then sent to the Wizard and stored for later analysis. The Wizard could also hear the participant through the online conferencing tool in which the study was being held, so as to increase accuracy of understanding the participant.

Speech Synthesis. After receiving a participant's message, the Wizard selected a response from the set of pre-defined statements on their own interface. For text-to-speech of the agent, CereProc was used (https://www.cereproc.com/), renowned for synthetic voices retaining naturalness and character — *"The CereVoice Engine SDK is the first free, commercial-grade, real-time speech synthesis system for academic research. It is fast, stable, and highly configurable, and is well suited to research into text-to-speech and dialogue applications."* CereProc voices allow for emotional synthesis control and each voice comes with vocal gestures such as laughs, coughs, and expressive interjections, e.g., "hmm", "ah", "yeah", "oh", etc. The voice chosen was Mairi - a child's voice with a Scottish-English accent.

5.2.3 Experimental Conditions

A between-subjects experimental design was used, with participants being randomly placed into one of three conditions: (1) Interjections, (2) Music, or (3) Control (no added auditory gestures).

The agent made both correct and incorrect statements about the topic being taught throughout the interaction. The statements and dialogue flow were developed through iterative pilot studies. Both correct and incorrect statements lead to moments where the agent replied with auditory gestures of positive or negative valence, respectively. Valence ranges from pleasant (positive) to unpleasant (negative) – examples of negative valence include sadness and fear, whereas happiness is an example of positive valence. There was an equal amount of both through the conversation: eight moments of positive, and eight moments of negative. The moments were selected by a native English speaker (first author) based on the fit within the conversational context. In each condition, the exact gesture was randomly selected from the set of either positive or negative ones, so as to increase variation across participants.

A total of fourteen positive and negative **interjections**, provided by the CereProc voice, were used in the study: seven with positive valence ("sigh", "hmm", "hmmm", "ah", "oh", "yay", "yeah") and seven with negative valence ("sigh", "ah", "oh", "ugh", "argh", "arr", "doh").

The **musical sounds** used in the study were taken from the validated set of auditory stimuli: The Musical Emotional Bursts (MEB) dataset [119]. It consists of 80 brief musical executions expressing positive and negative valence. The dataset contains both improvisations and imitations of emotional expressions made by violinists and clarinetists. The improvisations of the clarinetists were selected for the study as they were shown by Paquette et al. [119] to have a high recognition accuracy for both the positive and negative valences, resulting in a total of fourteen musical executions: seven positive and seven negative.

Each speech utterance that followed an auditory gesture (i.e., interjection or music) was also matched in valence (positive or negative) to the gesture, using the synthesis control provided by the CereProc voice. For example, if the gesture was positive, then the utterance following it, i.e., corresponding to it, was also synthesized with positive valence. This was done so that the agent expressed valence in all conditions including control, with the only difference being the added expressive auditory gestures in the two experimental conditions. Adjusting the gesture and speech utterance to match is also important as prior work indicates users can feel a disconnect between speech and sound if not matched appropriately (e.g., [7]). No specific personality of the agent was designed, however the agent made three off-task statements during the conversation: indicating enjoyment of the task, asking how they're doing so far on the task, and telling a rock joke, as well as asking some additional on-topic questions about rocks. These off-task moments were included as prior work indicates they can lead to more positive experiences in pedagogical agent interactions (e.g., [62]).

Table 5.1: Example conversation in the Interjections condition.

AGENT	What's the first question about metamorphic rocks?
USER	It says metamorphic rocks are made by blank.
AGENT	So metamorphic rocks are the ones that form from intense heat and pressure. Am I right?
USER	That's exactly right. You got it.
AGENT	***Yay***! There we go. [positive valence]
AGENT	Shall we move on to the next question?
USER	Yes good idea. The next question asks: what is foliation in metamorphic rocks caused by?
AGENT	What causes foliation? Well first of all, what does foliation mean? Is that when there are fossils in the rock?
USER	No it's not fossils. Foliation are the flat layers that form as the rock is squeezed by pressure.
AGENT	***Oh***... So it's not fossils, but layers in the rock. [negative valence]

5.2.4 Research Questions and Hypotheses

The study was guided by the following research question:

Q1: How do interjections and musical executions affect:

- the agent-learner relationship, i.e., perceived interaction quality and rapport,
- attitudes & behaviours, i.e., motivation, teaching behaviour, etc.,
- ability to recall material, and
- cognitive workload.

Based on cognitive load theory, some argue that pedagogical agents can impose extraneous cognitive load and be detrimental to learning outcomes (e.g., [36]). As this study involves dialogue with the addition of expressions that hold meaning, the impact on learners' cognitive load was investigated as well.

Given the results of prior work, it was hypothesized that separately, interjections and musical executions, would result in higher ratings of rapport and interaction quality, and that there would be an increase in the affective learning outcome – motivation – with increased rapport. In the context of learning-by-teaching with teachable agents, the increased motivation would lead to improvement in the cognitive learning outcome – recall.

5.2.5 Methodology

Procedure

The study began by a researcher and participant connecting via an online conferencing tool. From here, participants were given an information letter and consent form, and then asked to fill in a pre-study questionnaire that included demographics information (e.g., age, gender, experience with conversational agents, etc.) and a quiz to test their knowledge on the topic to be discussed during the interaction with the agent. The quiz included 14 questions (5 for each rock type - 1 question covers both metamorphic and sedimentary types) and were adapted from the Lumen Learning course. Three multiple choice answers were provided for each question. Some quiz questions were covered in the interaction with the agent directly, others were in the articles but not covered specifically by the agent's questions.

Participants were then introduced to the interface (Figure 5.1) on which they would complete the task. After signing in to the interface, participants were given 3 minutes to read through the articles and questions on the 'Rocks Worksheet'. Following this, the agent introduced themselves to the participant, asked them their name, and explained their task again briefly. Participants had this preliminary dialogue with the agent to reduce novelty effects before the actual experiment began. Once participants communicated to the agent they were ready to begin, the agent asked which worksheet questions on one of the three rock types (Sedimentary, Metamorphic, Igneous) the participant wanted to start with. This was done to provide participants with an opportunity to partially guide the interaction and to increase variation of the order in which the questions were covered. The interaction with the agent lasted around 20-30 minutes. Following the interaction, participants filled out a number of post-study questionnaires. In total, all questionnaires took approximately 15-20 minutes.

Measures

Human-Agent Relationship. To measure participants' perceived relationship with the agent, the rapport instrument used by other researchers [113, 59] was adapted - covering two rapport dimensions, Understanding (a sense of mutual understanding) and Emotional (a sense of emotional connection), and extended it with questions from [30] on Quality of Interaction. The 15 questions were presented with a five-point Likert scale from "strongly disagree" (1) to "strongly agree" (5), balanced for positive and negative responses. Pick-a-Mood, a cartoon-based pictorial instrument, was used to measure the perceived mood of the agent [40], and finally, participants were given a number of questions relating to experience of the teaching task and the agent as a student: “How much did you like teaching [the agent]?”, “Do you think you were good at teaching [the agent]?”, and “Do you think [the agent] was a good student?”. They rated their agreement with each question by selecting one of the following: not at all, a little, quite a lot, and very much, and were also provided the opportunity to give free-form answers.

Learning Outcomes. To evaluate cognitive learning outcomes, participants were asked to take the same quiz as prior to the interaction and changes in quiz score pre- and post-interaction were compared to measure recall of the material. Similar to Study II, the Academic Motivation Scale (AMS) [158] was used to investigate the affective learning outcome motivation. The wording of the questions vary slightly from the original to fit this study.

Cognitive Workload. The workload profile (WP; [155]) was used to investigate subjective cognitive workload and had participants rate the proportion of attentional resources used on multiple dimensions on a scale of 0 to 100: perceptual/central processing, response selection and execution, spatial and verbal processing, visual and auditory processing, and manual and speech output. The scores are then summed and for comparison, this sum is averaged.

Participants

The study received ethics clearance and 41 participants were recruited through mailing-lists and participant pools at a research-based institution. One participant did not consent to the use of their data, another did not comply with survey instructions, and another assumed from the beginning of the interaction that the study used the Wizard-of-Oz technique. Results reported are therefore based on 38 participants (21 women, 17 men; age range = 18-52 years; median = 24, SD = 5.75; 2 participants did not provide their age). All participants were volunteers and received a $15 gift card. Both native (61%) and non-native (39%) English speakers participated in the study, and varied in whether their post-secondary education related to STEM fields (84% STEM-related, 16% not), their highest completed or current degree (58% Bachelor’s, 32% Master’s, 8% Doctorate, 2% College credit), and their ethnicity (61% Asian, 18% White, 3% Aboriginal or Indigenous, 3% Asian and Native Hawaiian or other Pacific Islander, 3% Black or African American, 3% Asian and Hispanic, Latino or Spanish origin, 3% Middle Eastern or North African, 3% preferred not to disclose, and 3% self-described as Sikh Punjabi). Additionally, when asked to rate their level of interest and experience with conversational agents on a 7-point

Likert scale from 1 (not at all/never) to 7 (very interested/a lot), 8% were not at all or had very little interest in agents, 47% were moderately interested, and 45% were highly interested; and 26% had very little experience with agents, 42% had a moderate amount of experience, and 32% were highly experienced with conversational agents.

Table 5.2: Demographics Summary Table

	Control (n=11)	Music (n=13)	Interjections (n=14)	
age (years)	M=23.91 $\pm$3.14	M=24.08 $\pm$1.88	M=26.85 $\pm$8.91	F(2,33)=1.02 $p = 0.37$
gender	5man 6woman	5man 8woman	7man 7woman	$\chi^2(2, N = 38) = 0.37$ $p = 0.83$
native English	9yes 2no	6yes 7no	8yes 6no	$\chi^2(2, N = 38) = 3.28$ $p = 0.19$

5.3 Analysis

Both qualitative and quantitative data were collected from each participant. For the numerous measures the following analyses were carried out: ANOVA, Kruskal-Wallis, linear regression models and cumulative link model (CLM; models the cumulative probabilities of discrete ordinal categories [104, 3]), with condition and demographics (gender, native language, pre-interaction quiz score, etc.) as the independent factors, and Pick-a-Mood, Understanding and Emotional Rapport, Quality of Interaction, AMS, WP, and change in pre to post quiz score, as the dependent factors.

5.4 Results

5.4.1 Perception of Expressive Auditory Gestures and Learner-Agent Relationship

One-way ANOVA showed no significant difference between conditions on the measures of Quality of Interaction ($F(2, 35) = 1.54, p = 0.23$) and Understanding Rapport ($F(2, 35) = 1.63, p = 0.21$). Results of the free-form questions, analyzed with CLM, similarly showed no significant differences between conditions, and neither did the Pick-a-Mood pictorial self-report scale for the agent's mood and personality. However, condition was found to have an effect on the Emotional Rapport dimension ($F(2, 35) = 3.34, p = 0.05$), with Tukey's HSD showing that participants in the Interjections condition ($M = 4.48, SD = 0.43$) on average rated feeling significantly more Emotional Rapport with the agent than those in the Control condition ($M = 3.98, SD = 0.63$), at $p = 0.04$.

Participants' answers to the free-form questions suggest that both the Interjections (I) and Music (M) gestures were perceived as intended, e.g., *"She seemed embarrassed or proud of herself at times. She expressed these emotions through verbal noises such as an exclamation when she would get the answer right or wrong"* (I09); *"Cheerful and interested. She showed this by exaggerated "no's" when an answer was incorrect and excited when she got an answer correct"* (I13), and *"[the agent] expressed "happiness" with a happy music and sadness with a "sulky music" " (M10)*; *"[the agent] was very expressive ... there was happy or sad music as well whenever it tried to reciprocate its emotion"* (M13).

The majority of participants in the Interjections and Music conditions stated they enjoyed the experience and expressed perceptions that, *"[the agent] was very personable"* (I07), *"she was engaged"* (I09), *"I liked the smooth interaction and the flow of the conversation"* (M01), *"it was interactive and engaging"* (M04), and *"the jokes and songs were fun and interesting"* (M09). Overall, most participants enjoyed the addition of auditory gestures during the task, but for a few participants the gestures were considered *"a little scary"* (M05) and *"a bit creepy"* (I01). M14 explained, *"I would leave out the music. It takes away from the flow of the conversation"*. Similar to work that found children distinguish between 'creepy' sounds that express intent and non-threatening sounds that are spontaneous [170] – our results suggest similar sentiments in adults are possible. These statements also further support the importance of adjusting the gesture and speech to match, so as to maintain the flow of dialogue and enhance the experience.

5.4.2 Learning Outcomes

To investigate motivation (i.e., affective learning outcome) the AMS questionnaire was used which provided overall scores of intrinsic-, extrinsic-, and a-motivation, with each type being further distinguished into more specific motives. Analysis of each overall and subscale score was done using one-way ANOVA followed by Tukey's HSD.

Extrinsic Motivation. In terms of ratings of being extrinsically motivated to make the effort to teach the agent, there was a significant difference between conditions ($F(2, 35) = 6.19, p = 0.005$), with participants in the Interjections condition ($M = 4.26, SD = 1.29$) reporting on average significantly more extrinsic motivation than participants in the Control condition ($M = 2.82, SD = 1.08$), at $p = 0.004$. At a lower-level, while no significant differences were found in the EM - externally regulated or EM - introjected motives, there was a significant difference in the amount of EM - identified (behaviour motivated by the view that participation is important for personal growth) reported between conditions ($F(2, 35) = 9.83, p = 0.0004$), with participants in both the Interjections ($M = 5.50, SD = 0.96$, at $p = 0.0003$) and Music ($M = 4.67, SD = 1.12$, at $p = 0.03$) conditions rating their motivation in the task as EM - identified more highly than participants in the Control condition ($M = 3.39, SD = 1.47$). However, a linear model indicated that a higher pre-quiz score reduced the amount of reported EM - identified in both the Interjections ($\beta = -0.90, t(32) = -2.91, p = 0.007$) and Music ($\beta = -1.01, t(32) = -2.92, p = 0.006$) conditions compared to Control.

Intrinsic Motivation. With regards to intrinsic motivation, condition was found to have a significant effect on overall intrinsic motivation ($F(2, 35) = 8.44, p = 0.001$), with participants in the Interjections condition ($M = 5.84, SD = 1.02$) feeling more intrinsically motivated than participants in the Control condition ($M = 3.84, SD = 1.34$), at $p = 0.0007$, and some evidence of participants in the Music condition ($M = 5.03, SD = 1.29$) feeling more intrinsically motivated compared to Control as well, at $p = 0.06$. Condition was also found to have a significant effect in each of the sub-motives of intrinsic motivation:

- IM - toward accomplishment: engaging in actions for the pleasure and satisfaction experienced when trying to achieve something new or beyond one's limits ($F(2, 35) = 5.08, p = 0.01$). Participants in the Interjections condition ($M = 6.07, SD = 0.94$) reported feeling significantly more IM-toward accomplishment than those in the Control condition ($M = 4.23, SD = 1.94$), at $p = 0.009$.

- IM - to know: actions performed for the pleasure and satisfaction derived from learning, exploring, or trying to understand something new ($F(2, 35) = 6.48, p =$

0.004). Participants in the Interjections condition ($M = 6.09, SD = 0.87$) reported feeling significantly more IM-to know than those in the Control condition ($M = 4.43, SD = 1.37$), at $p = 0.003$.

- IM - to experience stimulation: motivation related to the experiencing of pleasurable sensations ($F(2, 35) = 8.50, p = 0.001$). Participants in both the Interjections ($M = 5.36, SD = 1.50$) and Music ($M = 4.51, SD = 1.60$) conditions reported feeling significantly more IM-to experience stimulation than those in the Control condition ($M = 2.79, SD = 1.59$), at $p = 0.0007$ and $p = 0.03$, respectively.

To investigate further the variables that impact the intrinsic motivation sub-motives, different linear models were used for each subcategory. Through step-wise selection, the resulting models show that participants' prior knowledge of the topic (i.e., higher pre-interaction quiz scores) significantly influenced feelings of IM-to know with participants in the Interjections ($\beta = -0.68, t(29) = -2.09, p = 0.05$) and Music ($\beta = -0.85, t(29) = -2.36, p = 0.03$) conditions reporting lower feelings of IM-to know than in the Control condition. Additionally, native English speakers reported feeling more IM-to experience stimulation in the Music condition than the Control condition ($\beta = 2.99, t(32) = 2.05, p = 0.05$).

A-motivation. Lastly, one-way ANOVA showed condition had no significant effect on feelings of a-motivation ($F(2, 35) = 0.83, p = 0.44$): Interjections - $M = 3.07, SD = 1.19$; Music - $M = 3.72, SD = 1.67$; Control - $M = 3.55, SD = 1.10$.

In terms of **recall of material** (i.e., cognitive learning outcome), quiz scores pre-interaction started relatively high in all conditions (out of 13) Interjections: $M = 7.07, SD = 2.16$; Music: $M = 8.08, SD = 1.50$; Control: $M = 7.91, SD = 1.22$, and on average increased post-interaction. Condition had no significant effect on change in quiz score from pre- to post-interaction, $F(2, 35) = 0.4, p = 0.67$.

5.4.3 Cognitive Workload

Two participants did not complete the Workload Profile questionnaire correctly and so their data was not included in the analysis for cognitive workload. In general, one-way ANOVA indicated that condition had no effect on total workload (summation of individual dimensions to provide an overall workload rating; $F(2, 33) = 0.44, p = 0.65$). Independent Kruskal-Wallis tests were also conducted to examine the differences in each individual workload dimension separately. No dimension showed a significant difference between conditions. However, after removing one outlier from the data set (participant's ratings were far below the bulk of the data), results of the Krukal-Wallis test ($\chi^2(2, N = 32) = 6.88$, $p = 0.03$) indicated differences on the *Speech* dimension: "How much attention was required for producing the speech response (e.g., engaging in a conversation, talk, answering questions)?". A post-hoc Dunn test showed that participants in the Music condition reported lower workload in this dimension than those in the Interjections condition ($Z = 2.56$, $p = 0.03$) and also a trend towards a lower rating than the Control condition ($Z = 1.76$, $p = 0.12$). As overall workload was not affected by interjections or music, and neither were the *Verbal*: "How much attention was required for verbal material (e.g., reading, processing linguistic material, listening to verbal conversations)?", and *Auditory*: "How much attention was required for executing the task based on the information auditorily received (ears)?" dimensions, these findings provide support that expressive auditory gestures in pedagogical agents can be implemented without being detrimental to cognitive resource availability.

5.4.4 Teaching Behaviour

Lastly, participants' responses to the agent's on-topic questions (i.e., teaching statements) were analyzed. The responses were categorized into three answer types: *non-informational* answers – answers providing no new information or knowledge – these included, acknowledgement: e.g., "yes", "uh-huh"; agreement: e.g., "that's it"; maybe/unknown: e.g., "something like that", "I don't know"; and rejection/disagreement: e.g., "that's not it", "no", *informational word-for-word* answers – answers containing information read word-for-word from the text provided, and *informational rephrase/reformulate* answers – answers providing information that was rephrased from the text or reformulated into a question. To compare the proportion of each answer type – calculated as the number of times an answer type was given divided by the total number of answers given – one-way ANOVAs and linear models with step-wise model selection were used. Condition was not found to have an effect on the proportion of giving non-informational, informational word-for-word, or information rephrase/reformulate answers.

Non-informational. Through step-wise linear regression, it was however found that participants who reported being more interested in conversational agents gave a smaller proportion of non-informational answers in the Interjections condition than in the Control ($\beta = -0.12, t(26) = -2.93, p = 0.007$) and Music ($\beta = -0.15, t(26) = -3.47, p = 0.002$) conditions, whereas participants with more prior knowledge of the topic being taught gave a higher proportion of non-informational answers in the Interjections condition compared to the Control ($\beta = 0.12, t(26) = 2.62, p = 0.01$) and Music ($\beta = 0.08, t(26) = 2.21, p = 0.04$) conditions. Lastly, native English speaking participants also gave a lower proportion of non-informational answers in the Interjections condition than the Music condition ($\beta = -0.41, t(26) = -3.25, p = 0.003$).

Informational word-for-word. Another linear model indicated that women gave a higher proportion of word-for-word answers in the Interjections condition compared to Music ($\beta = 0.14, t(32) = 2.07, p = 0.05$).

Informational rephrase/reformulate. Lastly, through step-wise linear regression, the resulting model indicated that participants that reported more interest in conversational agents and native English speaking participants, both gave a higher proportion of rephrase/reformulate answers in the Interjections condition than in the other two conditions: Interest: Music ($\beta = 0.24, t(29) = 4.69, p < 0.001$), Control ($\beta = 0.11, t(29) = 2.26, p = 0.03$); English: Music ($\beta = 0.60, t(29) = 4.12, p < 0.001$), Control ($\beta = 0.48, t(29) = 2.77, p = 0.01$). Conversely, participants reporting greater interest in conversational agents gave a lower proportion of rephrase/reformulate answers in the Music condition than in the Control ($\beta = -0.13, t(29) = -2.49, p = 0.02$).

5.5 Discussion

Given that prior work suggests the importance of the relationship between tutee and tutor for increasing learning gains for the tutor (e.g., [116]) and that the expression of emotional and/or cognitive states can strengthen the human-agent relationship (e.g., [1, 136]), the purpose of our study was to understand how learners (in the role of tutor) perceive a voice-based agent (in the role of tutee) that adds expressive auditory gestures, interjections or musical sounds, to synthetic speech to communicate, and what effects these expressions have on the interaction as well as learning outcomes. Our results indicate that adding expressive auditory gestures, especially in the form of interjections, significantly increases feelings of emotional rapport with the agent and enhances both intrinsic and extrinsic motivation towards the task. On-the-other-hand, brief expressive musical executions, although not resulting in increased feelings of rapport compared to the control, did lead to increases in certain dimensions of intrinsic and extrinsic motivation, and required less cognitive workload to provide speech responses for than when interjections were used.

Although significantly higher reported levels of emotional rapport and motivation were found in participants in the Interjections condition compared to the Control, there were no improvements to the cognitive learning outcome, recall. This may be due to participants hitting a ceiling on learning gains, as on average pre-quiz scores were relatively high. However, as research proposes that rapport between human and agent leads to learning gains, our findings provide evidence for a separation between rapport dimensions whereby developing understanding rapport is necessary for promoting cognitive learning outcomes, while feelings of emotional rapport enhance affective learning outcomes, such as motivation.

The lack of influence of interjections and musical sounds on improving understanding rapport or ratings of interaction quality may be indicative of the relatively short interaction time in the study – as rapport can take time to develop [151], or the way in which rapport was measured and defined (i.e., subjective measurement). It may also be a consequence of the agent being voice-based. Prior studies (e.g., [168]) have shown that mixing two modalities – sound and facial expressions – can lead to stronger emotion recognition than when presented individually. Indeed, two participants in the study mentioned that the agent could be given *"a face to make her more personable"* (I14) and to *"try bringing [the agent] on screen in the form of a figure to visualize better"* (M11).

The influence of the user's personality on the human-agent relationship is becoming increasingly apparent, e.g., [30, 31]. In fact, beyond condition effects, our results similarly elucidate how certain demographic characteristics can influence various of our measures. With regards to motivation, prior knowledge of the topic in the task had a negative influence on intrinsic motivation associated with wanting to learn, explore or understand something new, and extrinsic motivation related to personal growth, when participants interacted with the agents using either expressive auditory gesture. Our results also indicate that the amount of interest towards conversational agents that users expressed having prior to participating in the study had a significant positive impact on their teaching behaviour during the interaction with the Interjections agent. These findings are particularly important in the context of educational systems, and although overall the findings suggest enhanced motivation and feelings of emotional rapport with the addition of interjections, future studies can take into consideration personality characteristics to further our understanding of the effects.

5.5.1 Limitations

The role of the user in the conversation, as well as the conversational context, are important considerations when evaluating the use of expressive auditory gestures. In contrast to prior work, and the common role of voice-based agents i.e., as assistants, the role of the user in our study was an informer/teacher/tutor. This provides the context for the results in our study, and there is potential for future work to investigate whether the results generalize to other roles, contexts, and agents as well. Furthermore, future work can examine how valence (positive vs. negative) can further affect learning, e.g., only positive when learning about one topic and only negative when learning about another. Additionally, as others have done in more socially-oriented conversations, e.g., Cohn et al. [37], future studies can look to understand what effects filler words, or a combination of fillers and interjections, or even interjections and music, can have on the constructs measured in this study. Lastly, the study is limited by the relatively small sample size and short interaction time with the agent, making it difficult to generalize the results to larger or different populations. To handle small sample size, the analysis avoided including more covariates in the ANOVA and the simplest possible models were selected to explain the data in order to prevent overfitting.

5.6 Conclusion

This chapter presented a study in which expressive auditory interjections and musical sounds were added to the dialogue of a teachable pedagogical agent. Where previous studies have focused largely on expressions through facial, gestural, and verbal cues, this work explored the lesser researched modality of auditory gestures – and investigated both voice-dependent (interjection) and voice-independent (music) gestures. The results indicate that expressive interjections and musical executions can be used to convey moments of positive and negative affective state in a voice-based agent, and evidence for interjections leading to the feeling of a significantly stronger emotional connection with the agent. Interjections also appear to be particularly effective at positively influencing both intrinsic and extrinsic motivation in learners, without having detrimental effects to cognitive workload. However, in terms of recall of the material, again there was no significant effect of condition.

With interjections becoming more widely available in popular systems, e.g, *"speechcons"* in the Amazon Alexa, the results provide practical insights for voice-system designers in education as well as across other domains including healthcare, entertainment, and customer service, for example, in which building rapport and enhancing motivation can be of similar value.

Chapter 6

Conclusions

This thesis explores how expressions of curiosity, humour, and expressive auditory gestures in pedagogical agents can adapt the learning environment with the aim of influencing learner motivation. The work was guided by the following research questions:

> **RQ1:** What are the effects of having a pedagogical agent convey intrinsic motivation (i.e., curiosity) to learners using verbal expressions?
>
> **RQ2:** What are the effects of having a pedagogical agent express humour, using different humour styles?
>
> **RQ3:** What are the effects of adding expressive auditory gestures (i.e., interjections and musical executions) to the speech of a pedagogical agent?

To investigate these questions, three separate studies were conducted. In each case, the interest was in understanding: *Perception*, i.e., How do learners perceive/identify the implemented expression and how does the expression influence learners framing of the agent? — with respect to measures including: likeability, perceived intelligence, and rapport, *Emotion and Behaviour*, i.e., What are learners' emotional and behavioural responses (in regards to motivation) to the expression?, and *Learning*, i.e., Are there any gains in cognitive outcomes as a result of the expression?

6.1 Contributions and Impact

The findings from each study add to the existing body of work that proposes the beneficial effects that human-like social cues, can have on learners. In general, the results indicate that pedagogical agents using affective expressions – in particular, curiosity, humour, and expressive auditory gestures – can be designed in such a way that modifies learner perceptions of the agent and learning experience, enhances motivation, but has no effect on cognitive outcomes. The results challenge prior work that suggest associations between affective expressions and cognitive learning outcomes, but rather provides evidence that agents using affective expressions may lead to more positive affective states that allow a learner to focus and be motivated, without leading directly to learning. The findings further suggest that affective expressions mediate the relationship between learner's perceptions and behaviours, and perceived affective learning outcomes. The work provides insight into how these types of expression affect learners on multiple dimensions, which can be of use to those designing agents outside of education, as well as for a deeper understanding of how these expressions may be impacting motivation in educational contexts. This chapter aims to summarize the main findings, discuss insights, and provide design speculations for conversational pedagogical agents.

6.2 Summary of Findings

6.2.1 Curiosity

Study I presents a new structured game, *LinkIt!*, that can be used to design and study robot behaviour and human-robot interaction. In addition, a procedure for assessing curiosity through behaviour in a *Free Choice Curiosity Test* is revealed. The results show perception of curiosity in a curious learning companion robot, and the emotional and behavioural contagion effects on learners.

The findings support prior work related to simulating affective states in agents, to elicit the same affective state in the user. The curious verbal behaviour implemented is successful in creating a robot that behaves in a manner considered curious by learners, and the results also build off of prior work that indicated the possibility of behavioural curiosity contagion in children, after interacting with a curious pedagogical robot. There is further evidence that, with a different expression of curiosity (i.e., on-topic question-asking), both emotional *and* behavioural curiosity contagion can occur in university-level learners. The results also indicate that learners can mimic their dialogue partner within 30 minutes, and that in contrast to the findings of prior work, beyond mimicking the specific behaviour shown by the agent, on-topic question-asking appears to evoke other curiosity-driven verbal behaviours as well, similar to what is seen in human-human dialogue [145]. However, if the pedagogical agent provides *too* much information about it's curiosity, it's possible that emotional and/or behavioural contagion effects are hindered by the extra cognitive workload imposed to process the additional information, and for curiosity to persist, it appears that learners need to be able to resolve their knowledge gaps that arise during the interaction. Lastly, although it was hypothesized that the curiosity contagion would be accompanied by improved cognitive outcomes, there was no evidence for this.

6.2.2 Humour

Study II was guided by RQ2, and presented an investigation of humour within pedagogical agents, with a comparison of two humour styles. Little prior literature has been devoted to systematically exploring how humour, in the form of humour styles, may influence learners, especially in a teachable agent context.

Some researchers believe that humour should not be used by agents in on-topic conversation because it can be distracting, and although the humour was distracting for some participants, on the whole, on-topic, off-task humour was found to be entertaining and motivating. The results of Study II indicate that an affiliative humour style leads learners to put in more effort and feel more motivation, in particular – more externally-regulated, than when teaching an agent that has a self-defeating humour style. On the other hand, a self-defeating humour style leads learners to put in more effort, but has no influence on self-reports of intrinsic or extrinsic motivation. Additionally, in contrast to prior work on conversational agents outside of education, the use of humour through humour styles doesn't lead to increased task enjoyment in a teachable agent context, and the self-defeating humour style actually appears to make learners enjoy the experience less. There is also evidence for learners' characteristics interacting with the agents'. Self-defeating jokes appear to evoke empathy and increase effort in learners with more aggressive, self-defeating, or affiliative humour styles, improve the confidence in teaching of learners who have a self-defeating humour style themselves, but negatively impact the experience of learners with more of a self-enhancing humour style. An affiliative humour style increases effort in learners with an aggressive or affiliative humour style, but negatively impacts the effort of learners with a self-enhancing or self-defeating humour style. Similar to Study I, there is no evidence of an influence of humour styles on learning gains. This may be the result of the humour used not being perceived as funny, however this requires more research to be better understood.

6.2.3 Expressive Auditory Gestures

Study III focused on using expression of affect to strengthen the learner-agent relationship. Expressive interjections and a voice-independent version, musical executions, were implemented in a voice-based pedagogical agent. How these expressions are perceived in conversational agents and what effects they have on learners in educational contexts has been largely neglected by the prior literature.

Measures of Emotional & Understanding rapport, and interaction quality were used to gauge the learner-agent relationship, and there was strong evidence found for interjections leading to significantly more emotional rapport, accompanied by increases in both intrinsic and extrinsic motivation felt during the task. However, prior knowledge of the subject matter has a negative influence on intrinsic motivation associated with wanting to learn, explore or understand something new, and extrinsic motivation related to personal growth. Musical executions do not lead to any significant increases in reported interaction quality or rapport, but does result in increases in certain dimensions of intrinsic and extrinsic motivation. Considering our findings from Studies I and II indicating possible negative effects imposed by extraneous cognitive load (i.e., too much information or too many jokes), the effects of expressive auditory gestures on workload were also investigated in Study III, and suggest that musical executions require less workload to provide speech responses for than when interjections are used. Again, as in Studies I and II, Study III's results indicate that neither expressive auditory gesture influence cognitive learning outcomes.

6.3 Design Speculations & Challenges

The difficulty with developing learning companion/teachable agents is that much research on human-human dialogue focuses on behaviour of the teacher or tutor. In contrast, an understanding of how to best design pedagogical agents taking on the role of the tutee or equally-knowledgeable peer is lacking. Therefore, beyond the results found in each study that contribute more to the research areas that focus on curiosity, humour, and expressive auditory gestures separately, there were also commonalities across studies that can guide the design of affective expression in these types of agents in general. Moreover, a number of design challenges arise as a result of the role the agent plays as a learning companion or less knowledgeable peer, as well as the social nature of the interaction, and therefore may be faced by others aiming to develop this kind of pedagogical agent.

Perception of Affect. How affect is perceived is important as it can influence the learner's experience and their teaching behaviour. The teaching behaviour (in the case of learning companions and teachable agents) is where the learning happens for the human, and therefore this is an important design feature to be considered. Whether the expressions are perceived to convey intent or as spontaneous can affect believability of the agent and human-like behaviours can damage the learning experience, i.e., negative mood in the agent can decrease self-efficacy in the learner, and interjections can make the agent appear "scary" or "creepy" because it becomes "too human-like". Therefore, although human-like behaviours are desirable for improving understandability and naturalness of conversational agents, perception of affect can influence the believability of the agent, feelings of deception and therefore trust, as well as the uncanny valley effect, which are crucial in creating successful interactions. Designers need to be aware of how affective expressions are perceived so that unintended consequences are avoided.

Learner Characteristics. As with more and more work on pedagogical agents, it's becoming increasingly clear how important learner characteristics are. The results in this thesis particularly highlight how they can influence both motivational and behavioural responses in learning contexts. In the case of *humour*, the findings can be particularly useful in informing design decisions for learners with, for example, low school motivation, in which self-defeating and aggressive humour styles are more typical [135]. In terms of *interjections*, their effectiveness for influencing both intrinsic and extrinsic motivation is negatively impacted by prior knowledge, which is a significant factor in the design of teachable agents, where the human takes on the role of the more knowledgeable peer. But this also poses a challenge: how should learner characteristics be taken into account? Should learners be given questionnaires before interacting with pedagogical agents, or given the option to choose what type of agent to interact with based on their own intuition or preferences? Could an agent build a learner profile over time and adapt it's expressions, or will this cause the agent's attitudes and moods to appear random and inconsistent?

Learning Outcomes. Across all studies, there was no evidence for the expressions enhancing learning gains, however only one cognitive outcome was measured – recall. Therefore future study designs should aim to capture more than just one outcome. The studies results also indicate that a short amount of time is enough to influence affective outcomes like motivation, but more time appears to be needed to influence cognitive outcomes, especially recall of material. Moreover, in terms of companion and teachable agents, the act of teaching might need to be coupled with preparing to teach for affective expressions to have an impact on cognitive and longer-term learning. A better understanding of the cognitive processes involved in the learning-by-teaching phenomenon could be achieved through the design of pedagogical agents taking on the role of less-knowledgeable peer.

Agent Knowledge. How should learners' expectations of the agent's knowledge be handled, and how can a less or equally knowledgeable agent respond when learners are asking for more topical knowledge? Some learners expected the agent to be able to answer their questions when wanting to know more about the subject material, which may come from an abstraction of their experience with virtual assistants. When the agent was unable to provide more information, it lead to negative experiences, or even blocks to learning. A challenge for the design of such agents involves how to create the agent's knowledge to be appropriate for the learning context.

6.4 Ethical Considerations

As prior work has found, and as the findings from each study further support, affective expressions in conversational agents can influence users' behaviours, perceptions, and even affective states. This brings with it certain ethical implications that require careful consideration. The first, is the risk of bias, i.e., design bias in appearance and behaviour of the agent which for example can solidify specific gender roles, as well as algorithmic bias which can result in unfair outcomes, such as privileging one group of users over another. Designers need to be aware of these biases so that they can be avoided or mitigated.

Specifically in regard to the use of affective expressions in agents, since humans learn, find meaning, and develop their own identities through interactions with other social actors, if the agents display these social signals, there is a potential risk that users will ascribe intentions and feelings to the agent that are either not intended, or may lead users to make incorrect assumptions about the agent's capabilities. Depending on the role the agent takes opposite the user, or the type of user who is interacting (e.g., a child or member of other vulnerable population), this may pose more serious concerns. For example, users may believe that the agent misses them when they are gone, or are depressed when interacting with them, which may have consequences beyond what designers intend. Users may also come to depend on the agents or feel responsible for their emotional states. It is therefore crucial to conduct systematic studies of people's perceptions of, and reactions to, different agent modalities, functions, and features in order to develop responsible human-agent interactions. During the design process, experts on human social behaviour, such as teachers and psychologists, could also be involved in making decisions on what social behaviours should be included.

Expressions of affect may also be considered deceptive, since agents do not actually feel emotions, but users are usually not aware of how these types of systems work – they tend to anthropomorphise them. There is also the ethical concern of whether it is appropriate to deliberately design behaviour in an agent that is known to influence users' emotions or behaviours, and in doing so, potentially intruding on the user's autonomy. To combat these risks, designers need to avoid confusion between living and artificial systems, and be clear on their effects, through more transparency and explanation as to the capabilities.

6.5 Limitations & Open Questions

Affective expressions can be approached from multiple perspectives. This thesis focused on three specific types that were grounded in prior literature, and suggested to enhance learner motivation. Within each type: curiosity, humour, and expressive auditory gestures, there are a number of possible designs that can be implemented and studied. The studies presented in this thesis looked at specific implementations of each. Although this provides various insights for the design of these agents, there are certain limitations of the work, as well as numerous open questions that are left to be explored.

Design of Affect. As mentioned in the Design Speculations section, one interesting and under-explored aspect of the design of affective expressions is understanding how emotions are perceived by learners, i.e., as *unconscious & conscious* expressions, and what impacts they have on the interaction. Another feature of the design of affect involves *positive & negative* affective states. Many researchers and designers believe the expression of negative moods have no place in pedagogical agent scenarios, however it appears that negative moods can be useful for certain users, so when is the expression of negative emotions useful? for who? and in what context? And when is emotional contagion of negative affect not a problem? Furthermore, how can positive and negative affective states (or more specific ones) be used to enhance learning outcomes? For example, can pedagogical agent designers make use of the phenomenons of mood congruent and/or mood dependent recall to aid learners in enhancing learning gains? This can be particularly important as well for enhancing long-term memory outcomes. Lastly, emotions can be distinguished along various dimensions, including whether they are *basic & complex*. Basic emotions are associated with universally recognizable facial expressions and include happy, sad, angry, etc. Complex emotions on the other hand, have variable appearances and compositions, such as grief, regret, and curiosity. Complex emotions are considered to have a much larger cognitive component – making both their detection and expression more challenging. This opens various research directions including how to express complex emotions but also whether they can be detected as intended, and if not, within learning environments, do they impose too much cognitive workload?

Mechanism. The studies conducted in this thesis were based on hypotheses of adapting the learning environment in certain ways to enhance motivation, i.e., eliciting contagion effects, creating a positive learning experience, and strengthening the learner-agent relationship. However, whether these processes are the actual mechanisms behind the results is unclear. Other mechanisms may have been at play, such as initiating self-disclosure leading to a better understanding of internal state, enhancing self-efficacy beliefs in learners, making the agent more believable as a learner itself, or acting as a role-model for learners, e.g., question-asking. For example, interjections may have been particularly useful in the case where the agent is the tutee and learner the tutor, as it helped the human tutor better understand the agent tutee, i.e., forming a "mental model" of the agent, and tailoring their next pedagogical move to the agent's internal state as opposed to using more random teaching strategies. Investigating exactly the underlying mechanism behind the resulting motivational and behavioural changes remain open questions. A better understanding of the mechanism behind the resulting outcomes, would help further tease apart what behaviours should be implemented in pedagogical agents and why.

Agent Type. Different types of agent were not compared in the three studies – therefore many open questions remain in this space. With regards to affective expressions, and in particular curiosity, humour, and auditory gestures, the *form* the agent takes can be explored in more detail, i.e., text, voice, and embodied: how might the design of expressions change depending on the form and does form impact learning outcomes differently? Future work can also explore in more depth the *appearance* of the agent, both with regards to visual and auditory expressions, but also in terms of whether conversational agents and affective expressions need to be tied to human-like agents or is there space to manipulate various anthropomorphic or zoomporphic aspects of agents? *Modality* can also be explored, as different agents provide multiple modalities to provide information to the user through. Although the findings suggest that affect can be effectively communicated through text or voice-alone, how might the addition of other modalities further impact learning outcomes? In relation to curiosity and humour, how can curiosity and humour be expressed non-verbally and what impacts do they have when added to the studied verbal behaviours? This is of particular interest as sounds and vocalizations that lack semantics in natural spoken language but still provide a rich form of communication and expression (also known as semantic-free utterances; SFUs), are relatively under-investigated when used to express emotional state [169]. In fact, the design of SFUs is complex and there currently is no systematic process for generating SFUs to convey specific affective states [7]. There also exists little documentation on how to express curiosity through SFUs (most likely because of the multi-faceted nature of curiosity) and related to humour, laughter is suggested to be a very specific and distinct type of SFU that requires careful design for an effect on social bonding [153]. Therefore, future work could be devoted to exploring how to create curiosity expression through SFUs, what effects the addition of curiosity and humour SFUs (such as laughter) have on learners when added to the verbal expressions. Lastly, it appears that sensing the learner's affective state and adapting the agent's behaviour is not necessary for enhancing motivation, but the question remains whether a more *complex* agent that personalizes their behaviour to the learner can further impact outcomes? Personalization can involve timing, amount of information, but also taking into account learner characteristics. Each factor requires more systematic investigation before it can be implemented.

Interaction. The interaction in each study involved dyads, and therefore it remains to be explored how affective expressions can play off multiple human members in a group, but also multiple agents as well. How will the dynamics of these scenarios change perception and outcomes? How can learners be motivated when in groups – Will the same tactics apply?

Learning Outcomes. As mentioned previously, the studies in this thesis focused on one cognitive outcome, i.e., memorization of facts – or recall of information. Especially due to the social nature of conversational agents, future work would benefit from generating a better understanding of what types of cognitive outcomes are influenced by expressions of curiosity, humour, and auditory gestures, as well as affective expressions more generally. The studies were also designed to rely largely on self-reporting, especially in terms of measuring the affective outcomes of motivation. As suggested at the conclusion of Study I, to form a better understanding of the effects of affective expressions, objective measures can be included alongside the subjective ones. Furthermore, measurements were consistently taken at the conclusion of the interaction, while it may be of interest to understand temporal changes in affective states during the interaction to get a more detailed picture of the influence on learners as well as develop more complex agents. Lastly, in all studies, one topic was used for learning, i.e., rock formation. Although prior work has shown that social cues are also effective on learning outcomes with other topics, exactly how our results may generalize to other topics is yet to be determined.

6.6 Future Directions

Following the design speculations, challenges, limitations, and open questions, there are a number of future directions for research on pedagogical agents: 1) Using affective expressions with the goal of enhancing learner motivation shows promise as a pedagogical agent intervention, but also requires further attention to explore the link with other learning outcomes; 2) A crucial requirement for studies in this area is to conduct longer-term studies, especially as they relate to the expression of affect – how it relates to the agent's perceived personality, how/whether expressions should change over time, and how the relationship with the agent develops over time; 3) As the thesis findings indicate, the field can benefit from more systematic and finer-grained investigations into the design of agent features to develop a better understanding of human/learner perceptions and the impacts on emotional and behavioural responses; and 4) Researchers and designers of pedagogical agents can draw upon the wealth of knowledge provided by diverse research domains including social science, neuroscience, robotics, audiology, HCI, and HRI to inform study and design decisions, avoid reinventing the wheel, and consolidate findings to move the field forward. By considering these directions, future work can continue to enhance and further our understanding of pedagogical agent-based learning environments.

6.7 Final Thoughts

This thesis explored whether and how three different types of conversational agent can leverage expressions of curiosity, humour, and expressive auditory gestures – and in particular, foster motivation in learning environments. Beyond the implications for enhancing motivation, the work supports the viewpoint that for pedagogical agents to be successful they need more than just sophisticated technical capabilities, and that by acting socially and displaying expressive behaviours they can have emotional, cognitive, and behavioural impacts. The goal of the thesis was to combine theory and research spanning multiple domains to inform the development of effective pedagogical agents for learning environments that consider cognitive, emotional, and social elements, and the work contributes to our understanding of learner-agent interactions, what kinds of designs show potential, and how they can affect learners' perceptions and motivations.

References

[1] Jaime C Acosta and Nigel G Ward. Achieving rapport with turn-by-turn, user-responsive emotional coloring. *Speech Communication*, 53(9-10):1137–1148, 2011.

[2] Alan Agresti. *Categorical data analysis*. A Wiley-Interscience publication. Wiley, New York [u.a.], 1990.

[3] Alan Agresti. *Analysis of ordinal categorical data*, volume 656. John Wiley & Sons, 2010.

[4] Hua Ai, Rohit Kumar, Dong Nguyen, Amrut Nagasunder, and Carolyn P Rosé. Exploring the effectiveness of social capabilities and goal alignment in computer supported collaborative learning. In *International Conference on Intelligent Tutoring Systems*, pages 134–143. Springer, 2010.

[5] Esma Aïmeur, Hugo Dufort, Daniel Leibu, and Claude Frasson. Some justifications for the learning by disturbing strategy. In *Proceedings of the Eighth World Conference on Artificial Intelligence in Education*, pages 119–126. IOS Press, 1997.

[6] Safinah Ali, Tyler Moroso, and Cynthia Breazeal. Can children learn creativity from a social robot? In *Proceedings of the 2019 on Creativity and Cognition*, pages 359–368. 2019.

[7] Matthew P Aylett, Yolanda Vazquez-Alvarez, and Skaiste Butkute. Creating robot personality: Effects of mixing speech and semantic free utterances. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 110–112, 2020.

[8] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, 3(1):41–52, 2011.

[9] Sigal G Barsade. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative science quarterly*, 47(4):644–675, 2002.

[10] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81, 2009.

[11] Amy L Baylor and Soyoung Kim. Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior*, 25(2):450–457, 2009.

[12] Amy L Baylor and Yanghee Kim. Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, 15(2):95–115, 2005.

[13] Antoine Bechara, Hanna Damasio, Daniel Tranel, and Antonio R Damasio. Deciding advantageously before knowing the advantageous strategy. *Science*, 275(5304):1293–1295, 1997.

[14] Pascal Belin, Robert J Zatorre, and Pierre Ahad. Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, 13(1):17–26, 2002.

[15] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. Social robots for education: A review. *Science Robotics*, 3(21):eaat5954, 2018.

[16] Ronald A Berk. Student Ratings of 10 Strategies for Using Humor. *Journal on Excellence in College Teaching*, (January 1996), 1996.

[17] Ronald A Berk. *Professors are from Mars, Students are from Snickers*. Madison, WI: Mendota Press, 1998.

[18] Daniel E Berlyne. A theory of human curiosity. *British Journal of Psychology*, 45(3):180–191, 1954.

[19] Daniel E Berlyne. Conflict, arousal, and curiosity. 1960.

[20] Daniel E Berlyne. Curiosity and learning. *Motivation and Emotion*, 2(2):97–175, 1978.

[21] Timothy Bickmore and Justine Cassell. Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 396–403, 2001.

[22] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. Maintaining engagement in long-term interventions with relational agents. *Applied Artificial Intelligence*, 24(6):648–666, 2010.

[23] Timothy W Bickmore, Laura M Pfeifer Vardoulakis, and Daniel Schulman. Tinker: a relational agent museum guide. *Autonomous agents and multi-agent systems*, 27(2):254–276, 2013.

[24] Gautam Biswas, Thomas Katzlberger, John Bransford, Daniel Schwartz, et al. Extending intelligent learning environments with teachable agents to enhance learning. In *Artificial Intelligence in Education*, pages 389–397. Citeseer, 2001.

[25] Gautam Biswas, James R Segedy, and Kritya Bunchongchit. From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain. *International Journal of Artificial Intelligence in Education*, 26(1):350–364, 2016.

[26] Benjamin S Bloom. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, 13(6):4–16, 1984.

[27] Silvia Bogdan. Failed humour and its effects in conversation: a case study. *International Journal of Linguistics, Semiotics and literary Science*, 6(1):37–47, 2014.

[28] Carlos Salavera Bordás, Pablo Usan, Laurane Jarie, and Orosia Lucha. Sense of humor, affect, and personality. a study of university students. *Avances en Psicología Latinoamericana*, 36(1):83–91, 2018.

[29] Jennings Bryant, Paul W Comisky, Jon S Crane, and Dolf Zillmann. Relationship between college teachers' use of humor in the classroom and students' evaluations of their teachers. *Journal of educational psychology*, 72(4):511, 1980.

[30] Aleksandra Cerekovic, Oya Aran, and Daniel Gatica-Perez. How do you like your virtual agent?: Human-agent interaction experience through nonverbal features and personality traits. In *International Workshop on Human Behavior Understanding*, pages 1–15. Springer, 2014.

[31] Aleksandra Cerekovic, Oya Aran, and Daniel Gatica-Perez. Rapport with virtual agents: What do human social cues and personality explain? *IEEE Transactions on Affective Computing*, 8(3):382–395, 2016.

[32] Tanya L Chartrand and John A Bargh. The chameleon effect: the perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6):893, 1999.

[33] Kai-Yi Chin, Zeng-Wei Hong, and Yen-Lin Chen. Impact of using an educational robot-based learning system on students' motivation in elementary education. *IEEE Transactions on Learning Technologies*, 7(4):333–345, 2014.

[34] Jean M Civikly. Humor and the enjoyment of college teaching. *New Directions for Teaching and Learning*, 1986(26):61–70, 1986.

[35] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. What makes a good conversation? challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, Glasgow, Scotland, UK, 2019. ACM.

[36] Richard E Clark and Sunhee Choi. Five design principles for experiments on the effects of animated pedagogical agents. *Journal of Educational Computing Research*, 32(3):209–225, 2005.

[37] Michelle Cohn, Chun-Yen Chen, and Zhou Yu. A large-scale user study of an alexa prize chatbot: Effect of tts dynamism on perceived quality of social dialog. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, pages 293–306, 2019.

[38] Diana I Cordova and Mark R Lepper. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, 88(4):715, 1996.

[39] EL Deci and RM Ryan. Intrinsic motivation and self-determination in human behavior: Springer science & business media. 1985.

[40] Pieter M A Desmet, Martijn H Vastenburg, and Natalia Romero. Mood measurement with pick-a-mood: review of current methods and design of a pictorial self-report scale. *Journal of Design Research*, 14(3):241–279, 2016.

[41] Bernard J Dodge and Allison Rossett. Heuristics for humor in instruction. *Performance and Instruction*, 21(4):11, 1982.

[42] Stéphane Dupont, Hüseyin Çakmak, Will Curran, Thierry Dutoit, Jennifer Hofmann, Gary McKeown, Olivier Pietquin, Tracey Platt, Willibald Ruch, and Jérôme Urbain. Laughter research: a review of the ilhaire project. In *Toward Robotic Socially Believable Behaving Systems-Volume I*, pages 147–181. Springer, 2016.

[43] David Duran. Learning-by-teaching. evidence and implications as a pedagogical mechanism. *Innovations in Education and Teaching International*, 54(5):476–484, 2017.

[44] Pawel Dybala, Michal Ptaszynski, Shinsuke Higuchi, Rafal Rzepka, and Kenji Araki. Humor prevails!-implementing a joke generator into a conversational system. In *Australasian Joint Conference on Artificial Intelligence*, pages 214–225. Springer, 2008.

[45] Pawel Dybala, Michal Ptaszynski, Jacek Maciejewski, Mizuki Takahashi, Rafal Rzepka, and Kenji Araki. Multiagent system for joke generation: Humor and emotions combined in human-agent conversation. *Journal of Ambient Intelligence and Smart Environments*, 2(1):31–48, 2010.

[46] Pawel Dybala, Michal Ptaszynski, Rafal Rzepka, and Kenji Araki. Humorized computational intelligence towards user-adapted systems with a sense of humor. In *Workshops on Applications of Evolutionary Computation*, pages 452–461. Springer, 2009.

[47] Nancy Eisenberg, Robert Emde, Willard W Hartup, Lois Hoffman, Eleanor E Maccoby, Franz J Monks, Ross Parke, Michael Rutter, and Carolyn Zahn-Waxler. *Achievement and motivation: A social-developmental perspective*. Cambridge University Press, Cambridge, UK, 1992.

[48] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies*, 132:138–161, 2019.

[49] Laura Ferreri and Laura Verga. Benefits of music on verbal learning and memory: How and when does it work? *Music Perception: An Interdisciplinary Journal*, 34(2):167–182, 2016.

[50] Randy L Garner. Humor in pedagogy: How ha-ha can lead to aha! *College Teaching*, 54(1):177–180, 2006.

[51] Mirko Gelsomini, Hae Won, Jin Joo, and Cynthia Breazeal. Attentive Robot Listener Engages Children in Language Learning. In *International Conference on Social Robots in Therapy and Education (New Friends).*, 2016.

[52] Jacob W Getzels. The problem of interests: A reconsideration. *Supplementary Education Monographs*, 66(97):106, 1966.

[53] Sherri Gibson and Myron H Dembo. Teacher efficacy: A construct validation. *Journal of educational psychology*, 76(4):569, 1984.

[54] Robert Glenn. Brain research: Practical applications for the classroom. *Teaching for Excellence*, 21(6):1–2, 2002.

[55] Daniel Goleman. Emotional intelligence bantam books. *New York*, 1995.

[56] Goren Gordon, Cynthia Breazeal, and Susan Engel. Can children catch curiosity from a social robot? In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 91–98. ACM, 2015.

[57] Arthur C Graesser, Haiying Li, and Carol Forsyth. Learning by communicating in natural language with conversational agents. *Current Directions in Psychological Science*, 23(5):374–380, 2014.

[58] Arthur C Graesser and Natalie K Person. Question asking during tutoring. *American educational research journal*, 31(1):104–137, 1994.

[59] Ivan Gris Sepulveda. Physical engagement as a way to increase emotional rapport in interactions with embodied conversational agents. 2015.

[60] Matthias J Gruber, Bernard D Gelman, and Charan Ranganath. States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, 84(2):486–496, 2014.

[61] Charles R Gruner. Speaker ethos, self-disparaging humor, and perceived “sense of humor.". 1982.

[62] Agneta Gulz, Magnus Haake, and Annika Silvervarg. Extending a teachable agent with a social conversation module - Effects on student experiences and learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6738 LNAI:106–114, 2011.

[63] Habib Hamam. An e-learning interactive environment using animated pedagogical agent: Application to telecommunications courses. In *EdMedia+ Innovate Learning*, pages 4618–4621. Association for the Advancement of Computing in Education (AACE), 2005.

[64] William P Hampes. Humor and shyness: The relation between humor styles and shyness. *HUMOR*, 19(2):179 – 187, 2006.

[65] William P Hampes. The relation between humor styles and empathy. *Europe's Journal of Psychology*, 6(3):34–45, Aug. 2010.

[66] Elaine Hatfield, John T Cacioppo, and Richard L Rapson. Primitive emotional contagion. 1992.

[67] Deanna Hood, Séverin Lemaignan, and Pierre Dillenbourg. When Children Teach a Robot to Write: An Autonomous Teachable Humanoid Which Uses Simulated Handwriting. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2015.

[68] Jiaxiong Hu, Yun Huang, Xiaozhu Hu, and Yingqing Xu. *Enhancing the Perceived Emotional Intelligence of Conversational Agents through Acoustic Cues*. Association for Computing Machinery, New York, NY, USA, 2021.

[69] Eun-Sook Jee, Yong-Jeon Jeong, Chong Hui Kim, and Hisato Kobayashi. Sound design for emotion and intention expression of socially interactive robots. *Intelligent Service Robotics*, 3(3):199–206, 2010.

[70] Eun-Sook Jee, Chong Hui Kim, Soon-Young Park, and Kyung-Won Lee. Composition of musical sound expressing an emotion of robot based on musical factors. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 637–641. IEEE, 2007.

[71] Eun-Sook Jee, Soon-Young Park, Chong Hui Kim, and Hisato Kobayashi. Composition of musical sound to express robot's emotion with intensity and synchronized expression with robot's behavior. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 369–374. IEEE, 2009.

[72] Yuin Jeong, Juho Lee, and Younah Kang. Exploring effects of conversational fillers on user perception of conversational agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6. ACM, 2019.

[73] Jamie Jirout and David Klahr. Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review*, 32(2):125–160, 2012.

[74] Patrik N Juslin and Petri Laukka. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin*, 129(5):770, 2003.

[75] Patrik N Juslin and Petri Laukka. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of new music research*, 33(3):217–238, 2004.

[76] Peter H Kahn Jr, Jolina H Ruckert, Takayuki Kanda, Hiroshi Ishiguro, Heather E Gary, and Solace Shen. No joking aside: Using humor to establish sociality in hri. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, pages 188–189, Bielefeld, Germany, 2014. ACM/IEEE.

[77] Min Jeong Kang, Ming Hsu, Ian M Krajbich, George Loewenstein, Samuel M McClure, Joseph Tao-yi Wang, and Colin F Camerer. The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, 20(8):963–973, 2009.

[78] Todd B Kashdan, Matthew W Gallagher, Paul J Silvia, Beate P Winterstein, William E Breen, Daniel Terhar, and Michael F Steger. The curiosity and exploration inventory-ii: Development, factor structure, and psychometrics. *Journal of Research in Personality*, 43(6):987–998, 2009.

[79] Debra Korobkin. Humor in the classroom: Considerations and strategies. *College teaching*, 36(4):154–158, 1988.

[80] Kurt Kraiger, J Kevin Ford, and Eduardo Salas. Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of applied psychology*, 78(2):311, 1993.

[81] David R Krathwohl, Benjamin Samuel Bloom, and Bertram B Masia. *Taxonomy of educational objectives: The classification of educational goals; Handbook II: Affective domain.* David McKay Company, Incorporated, 1956.

[82] Nicholas A Kuiper and Nicola McHale. Humor styles as mediators between self-evaluative standards and psychological well-being. *The Journal of Psychology*, 143(4):359–376, July 2009.

[83] Philipp Kulms, Stefan Kopp, and Nicole C Krämer. Let's be serious and have a laugh: Can humor support cooperation with a virtual agent? In Timothy Bickmore, Stacy Marsella, and Candace Sidner, editors, *Intelligent Virtual Agents*, volume 8637 of *Lecture Notes in Computer Science*, pages 250–259, Cham, 2014. Springer International Publishing.

[84] H Chad Lane, Clara Cahill, Susan Foutz, Daniel Auerbach, Dan Noren, Catherine Lussenhop, and William Swartout. The effects of a pedagogical agent for informal science education on learner behaviors and self-efficacy. In *International Conference on Artificial Intelligence in Education*, pages 309–318. Springer, 2013.

[85] Edith Law, Parastoo Baghaei Ravari, Nalin Chhibber, Dana Kulic, Stephanie Lin, Kevin D Pantasdo, Jessy Ceha, Sangho Suh, and Nicole Dillen. Curiosity notebook: A platform for learning by teaching conversational agents. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, page 1–9, New York, NY, USA, 2020. Association for Computing Machinery.

[86] Edith Law, Vicky Cai, Qi Feng Liu, Sajin Sasy, Joslin Goh, Alex Blidaru, and Dana Kulić. A wizard-of-oz study of curiosity in human-robot interaction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 607–614. IEEE, 2017.

[87] Jong-Eun Roselyn Lee, Clifford Nass, Scott Brenner Brave, Yasunori Morishima, Hiroshi Nakajima, and Ryota Yamada. The case for caring colearners: The effects of a computer-mediated colearner agent on trust and learning. *Journal of Communication*, 57(2):183–204, 2007.

[88] SeoYoung Lee and Junho Choi. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, 103:95–105, 2017.

[89] Krittaya Leelawong and Gautam Biswas. Designing Learning by Teaching Agents The Betty's Brain System. *International Journal Artificial Intelligence in Education*, 18(3):181–208, 2008.

[90] James C Lester, Sharolyn A Converse, Susan E Kahler, S Todd Barlow, Brian A Stone, and Ravinder S Bhogal. The persona effect: affective impact of animated pedagogical agents. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pages 359–366, 1997.

[91] James C Lester, Jennifer L Voerman, Stuart G Towns, and Charles B Callaway. Cosmo: A life-like animated pedagogical agent with deictic believability. In *Proceedings of the International Joint Conference on Artificial Intelligence Workshop on Animated Interface Agents: Making them Intelligent*, 1997.

[92] Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 34, 2012.

[93] Tze Wei Liew, Nor Azan Mat Zin, and Noraidah Sahari. Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. *Human-centric Computing and Information Sciences*, 7(1):9, 2017.

[94] Brian Y Lim, Anind K Dey, and Daniel Avrahami. Why and why not explanations improve the intelligibility of context-aware intelligent systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2119–2128. ACM, 2009.

[95] Jordan A Litman. Interest and deprivation factors of epistemic curiosity. *Personality and individual differences*, 44(7):1585–1595, 2008.

[96] George Loewenstein. The Psychology of Curiosity: A Review and Reinterpretation. *Psychological Bulletin*, 116(1):75–98, 1994.

[97] Nichola Lubold, Erin Walker, and Heather Pon-Barry. Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 255–262. IEEE, 2016.

[98] Heidy Maldonado, Jong-Eun Roselyn Lee, Scott Brave, Cliff Nass, Hiroshi Nakajima, Ryota Yamada, Kimihiko Iwamura, and Yasunori Morishima. We learn better together: enhancing elearning with emotional characters. In *Proceedings of the 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!*, pages 408–417. International Society of the Learning Sciences, 2005.

[99] Nicola Mammarella, Beth Fairfield, and Cesare Cornoldi. Does music enhance cognitive performance in healthy older adults? the vivaldi effect. *Aging clinical and experimental research*, 19(5):394–399, 2007.

[100] Rod A Martin and Thomas Ford. *The psychology of humor: An integrative approach.* Academic press, Elsevier, 2018.

[101] Rod A Martin, Patricia Puhlik-Doris, Gwen Larsen, Jeanette Gray, and Kelly Weir. Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality*, 37:48–75, 2003.

[102] David Maulsby, Saul Greenberg, and Richard Mander. Prototyping an intelligent agent through wizard of oz. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 277–284, 1993.

[103] Richard E Mayer. Principles of multimedia learning based on social cues: Personalization, voice, and image principles. 2005.

[104] Peter McCullagh and JA Nelder. Generalized linear models ii, 1989.

[105] Michael Frederick McTear, Zoraida Callejas, and David Griol. *The conversational interface*, volume 6. Springer, 2016.

[106] Roxana Moreno, Richard E Mayer, Hiller A Spires, and James C Lester. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and instruction*, 19(2):177–213, 2001.

[107] John Morkes, Hadyn K Kernal, and Clifford Nass. Effects of Humor in Task-Oriented Human- Computer Interaction and Computer-Mediated Communication: A Direct Test of SRCT Theory. *Human Computer Interaction*, 14(4):395–435, 1999.

[108] Bonnie M Muir. Trust in automation: Part i. theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, 37(11):1905–1922, 1994.

[109] Michael Neff, Yingying Wang, Rob Abbott, and Marilyn Walker. Evaluating the effect of gesture and language on personality perception in conversational agents. In *International Conference on Intelligent Virtual Agents*, pages 222–235. Springer, 2010.

[110] Andreea Niculescu, Betsy van Dijk, Anton Nijholt, Haizhou Li, and Swee Lan See. Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy. *International Journal of Social Robotics*, 5(2):171–191, 2013.

[111] Andreea I Niculescu and Rafael E Banchs. Strategies to cope with errors in human-machine spoken interactions: using chatbots as back-off mechanism for task-oriented dialogues. In *Proceedings of ERRARE, Errors by Humans and Machine in multimedia, multimodal and multilingual data processing*, Sinaia, Romania, 2015. Editura Academiei Romane.

[112] Marret K Noordewier and Eric van Dijk. Curiosity and time: from not knowing to almost knowing. *Cognition and Emotion*, 31(3):411–421, 2017.

[113] David Novick and Iván Gris. Building rapport between human and eca: A pilot study. In *International Conference on Human-Computer Interaction*, pages 472–480. Springer, 2014.

[114] Benjamin D Nye, Arthur C Graesser, and Xiangen Hu. Autotutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, 24(4):427–469, 2014.

[115] Amy Ogan, Samantha Finkelstein, Elijah Mayfield, Claudia D'Adamo, Noboru Matsuda, and Justine Cassell. "Oh dear stacy!": social interaction, elaboration, and learning with teachable agents. In *Proceedings of the 2012 ACM Annual conference on Human Factors in Computing Systems*, pages 39–48, 2012.

[116] Amy Ogan, Samantha Finkelstein, Erin Walker, Ryan Carlson, and Justine Cassell. Rudeness and rapport: Insults and learning gains in peer tutoring. In *International Conference on Intelligent Tutoring Systems*, pages 11–21. Springer, 2012.

[117] Stefan Olafsson, Teresa K O'Leary, and Timothy W Bickmore. Motivating health behavior change with humorous virtual agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, IVA '20, New York, NY, USA, 2020. Association for Computing Machinery.

[118] Pierre-Yves Oudeyer, Jacqueline Gottlieb, and Manuel Lopes. Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. In *Progress in Brain Research*, volume 229, pages 257–284. Elsevier, 2016.

[119] Sébastien Paquette, Isabelle Peretz, and Pascal Belin. The "musical emotional bursts": a validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in Psychology*, 4:509, 2013.

[120] Bhargavi Paranjape, Yubin Ge, Zhen Bai, Jessica Hammer, and Justine Cassell. Towards automatic generation of peer-targeted science talk in curiosity-evoking virtual

agent. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 71–78, 2018.

[121] Hae Won Park, Rinat Rosenberg-Kima, Maor Rosenberg, Goren Gordon, and Cynthia Breazeal. Growing Growth Mindset with a Social Robot Peer. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, pages 137–145, 2017.

[122] Dybala Pawel, Ptaszynski Michal, Rzepka Rafal, and Araki Kenji. Humoroids: Conversational Agents That Induce Positive Emotions with Humor (Extended Abstract). In *AAMAS '09 Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, number 2, pages 1171–1172, 2009.

[123] Project for Education Research That Scales (PERTS). Growth mindset assessment, 2015.

[124] Laura M Pfeifer and Timothy Bickmore. Should agents speak like, um, humans? the use of conversational fillers by virtual agents. In *International Workshop on Intelligent Virtual Agents*, pages 460–466. Springer, 2009.

[125] John P Powell and Lee W Andresen. Humour and teaching in higher education. *Studies in Higher Education*, 10(1):79–90, 1985.

[126] Aditi Ramachandran, Chien-Ming Huang, Edward Gartland, and Brian Scassellati. Thinking aloud with a tutoring robot to enhance learning. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 59–68. ACM, 2018.

[127] Taffy E Raphael and P David Pearson. Increasing students' awareness of sources of information for answering questions. *American Educational Research Journal*, 22(2):217–235, 1985.

[128] Byron Reeves and Clifford Ivar Nass. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press, 1996.

[129] Thomas G Reio Jr. and Albert Wiswell. Field investigation of the relationship among adult curiosity, workplace learning, and job performance. *Human Resource Development Quarterly*, 11(1):5–30, 2000.

[130] Katerina Rnic, David J A Dozois, and Rod A Martin. Cognitive distortions, humor styles, and depression. *Europe's Journal of Psychology*, 12(3):348–362, August 2016.

[131] L R Roehler and D J Cantlon. Scaffolding: A powerful tool in social constructivist classrooms. *Scaffolding student learning: Instructional approaches and issues*, pages 6–42, 1997.

[132] Rod D Roscoe and Michelene TH Chi. Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, 77(4):534–574, 2007.

[133] John A Ross, Anne Hogaboam-Gray, and Peter Gray. Prior student achievement, collaborative school processes, and collective teacher efficacy. *Leadership and Policy in Schools*, 3(3):163–188, September 2004.

[134] Richard M Ryan and Edward L Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1):54–67, 2000.

[135] Vassilis Saroglou and Christel Scariot. Humor styles questionnaire: personality and educational correlates in belgian high school and college students. *European Journal of Personality*, 16(1):43–54, January 2002.

[136] Yuko Sasa and Véronique Auberge. Socio-affective interactions between a companion robot and elderly in a smart home context: prosody as the main vector of the "socio-affective glue". *SpeechProsody 2014*, 2014.

[137] E Glenn Schellenberg and Michael W Weiss. Music and cognitive abilities. 2013.

[138] Dale H Schunk. Self-efficacy and academic motivation. *Educational psychologist*, 26(3-4):207–231, 1991.

[139] Dale H Schunk. Coming to terms with motivation constructs. *Contemporary Educational Psychology*, 25(1):116–119, 2000.

[140] Dale H Schunk, Paul R Pintrich, and Judith L Meece. Motivation in education: Theory, research, and applications. 2008.

[141] Prachi E Shah, Heidi M Weeks, Blair Richards, and Niko Kaciroti. Early childhood curiosity and kindergarten reading and math academic achievement. *Pediatric Research*, April 2018.

[142] Masahiro Shiomi, Takayuki Kanda, Iris Howley, Kotaro Hayashi, and Norihiro Hagita. Can a social robot stimulate science curiosity in classrooms? *International Journal of Social Robotics*, 7(5):641–652, 2015.

[143] Toshiyuki Shiwa, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. How quickly should communication robots respond? In *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 153–160. IEEE, 2008.

[144] Elaine Short, Justin Hart, Michelle Vu, and Brian Scassellati. No fair!! an interaction with a cheating robot. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 219–226. IEEE, 2010.

[145] Tanmay Sinha, Zhen Bai, and Justine Cassell. A new theoretical framework for curiosity for learning in social contexts. In *European Conference on Technology Enhanced Learning*, pages 254–269. Springer, 2017.

[146] Tanmay Sinha, Ran Zhao, and Justine Cassell. Exploring socio-cognitive effects of conversational strategy congruence in peer tutoring. In *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And InfLuence*, INTERPERSONAL '15, pages 5–12, New York, NY, USA, 2015. ACM.

[147] Deborah A Small and Nicole M Verrochi. The face of need: Facial emotion expression on charity advertisements. *Journal of marketing research*, 46(6):777–787, 2009.

[148] Aimee E Stahl and Lisa Feigenson. Observing the unexpected enhances infants' learning and exploration. *Science*, 348(6230):91–94, 2015.

[149] Ron Tamborini and Dolf Zillmann. College students' perception of lecturers using humor. *Perceptual and Motor Skills*, 52(2):427–432, 1981.

[150] Pat M Taylor. An experimental study of humor and ethos. *Southern Speech Communication Journal*, 39(4):359–366, June 1974.

[151] Linda Tickle-Degnen and Robert Rosenthal. The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4):285–293, 1990.

[152] Margaret L Traeger, Sarah Strohkorb Sebo, Malte Jung, Brian Scassellati, and Nicholas A Christakis. Vulnerable robots positively shape human conversational dynamics in a human–robot team. *Proceedings of the National Academy of Sciences*, 117(12):6370–6375, 2020.

[153] Jürgen Trouvain and Marc Schröder. How (not) to add laughter to synthetic speech. In *Tutorial and Research Workshop on Affective Dialogue Systems*, pages 229–232. Springer, 2004.

[154] Jason Tsai, Emma Bowring, Stacy Marsella, Wendy Wood, and Milind Tambe. A study of emotional contagion with virtual characters. In Yukiko Nakano, Michael Neff, Ana Paiva, and Marilyn Walker, editors, *Intelligent Virtual Agents. IVA 2012. Lecture Notes in Computer Science*, volume 7502, pages 81–88. Springer, Berlin, Heidelberg, 2012.

[155] Pamela S Tsang and Velma L Velazquez. Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39(3):358–381, 1996.

[156] Markku Turunen, Jaakko Hakulinen, Cameron Smith, Daniel Charlton, Li Zhang, and Marc Cavazza. Physically embodied conversational agents as health and fitness companions. In *Ninth Annual Conference of the International Speech Communication Association*, 2008.

[157] Robert J Vallerand, Marc R Blais, Nathalie M Brière, and Luc G Pelletier. Construction et validation de l'échelle de motivation en éducation (eme). *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 21(3):323, 1989.

[158] Robert J Vallerand, Luc G Pelletier, Marc R Blais, Nathalie M Briere, Caroline Senecal, and Evelyne F Vallieres. The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement*, 52(4):1003–1017, 1992.

[159] Symeon P Vlachopoulos and Costas I Karageorghis. Interaction of external, introjected, and identified regulation with intrinsic motivation in exercise: relationships with exercise enjoyment. *Journal of Applied Biobehavioral Research*, 10(2):113–132, 2005.

[160] Lev Semenovich Vygotsky. In Michael Cole, Vera John-Steiner, Sylvia Scribner, and Ellen Souberman, editors, *Mind in society: The development of higher psychological processes*. Harvard university press, Cambridge, MA, 1978.

[161] Erin Walker, Nikol Rummel, and Kenneth R Koedinger. Integrating collaboration and intelligent tutoring data in the evaluation of a reciprocal peer tutoring environment. *Research and Practice in Technology Enhanced Learning*, 4(03):221–251, 2009.

[162] Chin-Yeh Wang, Shu-Yu Ke, Hui-Chun Chuang, He-Yun Tseng, and Gwo-Dong Chen. E-learning system design with humor and empathy interaction by virtual human to improve students' learning. In *Proceedings of the 18th International Conference on Computers in Education. Putrajaya, Malaysia: Asia-Pacific Society for Computers in Education.(ICCE)*, pages 615–622, 2010.

[163] Ning Wang, W Lewis Johnson, Richard E Mayer, Paola Rizzo, Erin Shaw, and Heather Collins. The politeness effect: Pedagogical agents and learning outcomes. *International journal of human-computer studies*, 66(2):98–112, 2008.

[164] Noel Wigdor, Joachim de Greeff, Rosemarijn Looije, and Mark A Neerincx. How to improve human-robot interaction with conversational fillers. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 219–224. IEEE, 2016.

[165] Q Wu, C Miao, and Z Shen. A Curious Learning Companion in Virtual Learning Environment. In *IEEE World Congress on Computational Intelligence (WCCI)*, 2012.

[166] Qiong Wu, Chunyan Miao, and Cyril Leung. Modeling curiosity in virtual companions to improve human learners' learning experience. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2015 IEEE/WIC/ACM International Conference on*, volume 2, pages 59–66. IEEE, 2015.

[167] Qiong Wu, Chunyan Miao, Xuehong Tao, and MG Helander. A curious companion for elderly gamers. In *Network of Ergonomics Societies Conference (SEANES), 2012 Southeast Asian*, pages 1–5. Ieee, 2012.

[168] Selma Yilmazyildiz, David Henderickx, Bram Vanderborght, Werner Verhelst, Eric Soetens, and Dirk Lefeber. Multi-modal emotion expression for affective human-robot interaction. In *Proceedings of the Workshop on Affective Social Speech Signals (WASSS 2013), Grenoble, France*, 2013.

[169] Selma Yilmazyildiz, Robin Read, Tony Belpeame, and Werner Verhelst. Review of semantic-free utterances in social human–robot interaction. *International Journal of Human-Computer Interaction*, 32(1):63–85, 2016.

[170] Jason C Yip, Kiley Sobel, Xin Gao, Allison Marie Hishikawa, Alexis Lim, Laura Meng, Romaine Flor Ofiana, Justin Park, and Alexis Hiniker. Laughing is scary, but farting is cute: A conceptual model of children's perspectives of creepy technologies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2019.

[171] Avner Ziv. The influence of humorous atmosphere on divergent thinking. *Contemporary Educational Psychology*, 8(1):68–75, 1983.

[172] Avner Ziv. Teaching and learning with humor: Experiment and replication. *The Journal of Experimental Education*, 57(1):4–15, 1988.

[173] Avner Ziv, Eli Gorenstein, and Anat Moris. Adolescents' evaluation of teachers using disparaging humour. *Educational Psychology*, 6(1):37–44, 1986.

APPENDICES

Appendix A

Study I - Curiosity

A.1 Questionnaires

This section includes the set of questionnaires used in Chapter 3 that are not otherwise publicly available.

A.1.1 Demographics

1. **Gender**

   ❍ male

   ❍ female

   ❍ other

   ❍ prefer not to say

2. **Age**

   ____

3. **What department are you in?**

   ____

4. **What degree program are you currently in?**

   ❍ Undergraduate

   ❍ Master

   ❍ PhD

   ❍ Other:

   ____

5. **Are you a native English speaker?**

   ❍ Yes

   ❍ No

6. **On a scale of 1-7, how interested are you in rocks?**

   ❍ 1 - Not interested at all

   ❍ 2

   ❍ 3

   ❍ 4

   ❍ 5

   ❍ 6

   ❍ 7 - Very interested

7. **On a scale of 1-7, how much do you know about rocks?**

   ❍ 1 - I know nothing

   ❍ 2

   ❍ 3

   ❍ 4

   ❍ 5

   ❍ 6

   ❍ 7 - I am an expert

8. **How often have you interacted with robots?**
   - ❍ 1 - Never
   - ❍ 2
   - ❍ 3
   - ❍ 4
   - ❍ 5
   - ❍ 6
   - ❍ 7 - A lot
9. **If you have interacted with robots before, what type of robots were they?**

   ____

10. **How interested are you in robots?**
    - ❍ 1 - Not at all
    - ❍ 2
    - ❍ 3
    - ❍ 4
    - ❍ 5
    - ❍ 6
    - ❍ 7 - Very interested

A.1.2 Knowledge Quiz

1. **Fossils are found only in**
   - ❍ Sedimentary rocks
   - ❍ Intrusive igneous rocks
   - ❍ Extrusive igneous rocks
   - ❍ Metamorphic rocks
   - ❍ I don't know

2. **Metamorphic rocks are formed by**

   - ❍ Heat and pressure
   - ❍ Cooling and hardening
   - ❍ The breaking down of other rocks
   - ❍ None of the above
   - ❍ I don't know

3. **Rocks that are formed from volcanoes are**

   - ❍ Granite
   - ❍ Igneous
   - ❍ Sedimentary
   - ❍ Sandstone
   - ❍ I don't know

4. **Sedimentary rocks are formed by**

   - ❍ Cooling and hardening
   - ❍ Deposition and cementation
   - ❍ Heat and pressure
   - ❍ Melting and cooling
   - ❍ I don't know

5. **Why do holes form in some rocks?**

   - ❍ Gas bubbles got trapped in the rocks
   - ❍ The rock cooled under water
   - ❍ The rock hardened underground
   - ❍ Because of pressure on the rock
   - ❍ I don't know

6. **Why do some rocks have a glassy surface?**

   ❍ Extreme pressure
   ❍ Rapid cooling
   ❍ High temperatures
   ❍ Melting of sand particles
   ❍ I don't know

7. **Which rock types can have layers?**

   ❍ Sedimentary & Metamorphic
   ❍ Metamorphic & Igneous
   ❍ Igneous & Sedimentary
   ❍ Only sedimentary
   ❍ I don't know

A.1.3 Feature & Type Quiz

Try to identify the features each rock has and the type that it is.

	Crystals	Layers	Glassy surface	Holes	Fossils	Sand	Pebbles (rounded rocks)	None of the above	I don't know	Igneous	Metamorphic	Sedimentary	I don't know
A	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
1	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
B	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
2	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
C	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
3	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
D	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
4	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
E	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
5	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
F	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
6	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
G	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
7	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
H	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
8	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
I	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
9	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
J	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
10	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
K	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
11	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
L	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○
12	☐	☐	☐	☐	☐	☐	☐	☐	☐	○	○	○	○

A.2 Interview Questions

This section covers the Interview questions used in Chapter 3.

A.2.1 Session 1

Questions about free period

1. What did you do during the free period and why?
    (a) What was going through your mind...
    (b) Which rock did you pick first? Why did you pick it first?
    (c) What did you think of NAO's response?

Questions about NAO and participant

1. How would you describe your experience with NAO?
    (a) Tell me more about that...
    (b) Can you give me a specific example of when that happened during the game?
2. On a scale of 1-7, how human-like did you find NAO?
    (a) What kinds of human-like characteristics would you say NAO has?
3. Would you say NAO has any particular personality/characteristics?
4. Would you say NAO is enthusiastic or not enthusiastic about rocks?
    (a) If yes, on a scale of 1 to 7, rate how enthusiastic you think NAO was about rocks.
        i. What did NAO say or do that made you think they were enthusiastic?
5. Would you say you are enthusiastic or not enthusiastic about rocks?
    (a) If yes, on a scale of 1 to 7, rate how enthusiastic you think you are about rocks.
6. Would you say NAO was engaged or not engaged in the game?

(a) If yes, on a scale of 1 to 7, rate how engaged you think NAO was in the game.
   i. What did NAO say or do that made you think they were engaged?

7. Would you say you were engaged or not engaged in the game?
   (a) If yes, on a scale of 1 to 7, how engaging did you find the experience with NAO?
      i. Why or why not?
      ii. Can you give me a specific example?

8. Would you say NAO is curious or not curious about rocks?
   (a) If yes, on a scale of 1 to 7, rate how curious you think NAO was about rocks.
      i. Why?

9. Would you say you are curious or not curious about rocks?
   (a) If yes, on a scale of 1 to 7, how curious are you?
      i. Has your level of curiosity changed after today?

10. Would you say NAO is curious or not curious about you?
   (a) If yes, on a scale of 1 to 7.

11. Are you curious or not curious about NAO?
   (a) If yes, on a scale of 1 to 7.
      i. Why or why not.

12. On a scale of 1 to 7, how much do you think NAO learned about rocks? (1 being nothing at all, and 7 meaning they are an expert now)

13. Would you say NAO is knowledgeable or not knowledgeable about rocks?
   (a) If yes, on a scale of 1 to 7, how knowledgeable on rocks do you think NAO is?
      i. What did NAO say or do that made you think they were knowledgeable?

14. On a scale of 1 to 7, how much do you think you've learned about rocks? (1 being nothing at all, 7 meaning you feel like an expert now)

15. What did you learn about rocks? Be specific.
   (a) Tell me as many things you can think of that you learned about rocks.

Questions about the game

1. Did you have a strategy during the game?
    (a) Were there certain rocks you were focusing more on during the game?
2. If given the opportunity, would you play the same game again with NAO? (yes or no)
    (a) Why or why not?
    (b) If you are given a choice to play this game again with a human or NAO, which would you choose? Why?
3. For the future, how do you think we could improve the game?

General Questions

1. If you could redesign all of this, what would you change about NAO?
2. On a scale of 1 to 7, how autonomous do you think NAO is? (1 being not autonomous at all, and 7 being fully autonomous)

A.2.2 Session 2

1. Over the past week, did you have any questions about rocks?
2. Over the past week, did you look up any information related to rocks?
3. On a scale of 1 to 7, how curious would you say you are about rocks? (1 being not curious at all, and 7 is extremely curious)
4. Do you remember anything specifically from your time with NAO last week? Anything that he did or said for example.

Appendix B

Study II - Humour

B.1 Questionnaires

This section includes the set of questionnaires used in Chapter 4 that are not otherwise publicly available.

B.1.1 Demographics

1. **What is your gender**

   ❍ man

   ❍ woman

   ❍ non-binary

   ❍ prefer not to disclose

   ❍ Other:

   ____

2. **What is your age?**

   ____

3. **If you are currently enrolled as a student, what degree will you receive upon graduation? If you are not currently a student, what is the highest degree you have completed?**

   ❍ High school graduate, diploma or the equivalent (e.g., GED)

   ❍ Some college credit, no degree

   ❍ Trade/technical/vocational training

   ❍ Associate degree

   ❍ Bachelor's degree

   ❍ Master's degree

   ❍ Professional degree

   ❍ Doctorate degree

   ❍ prefer not to disclose

   ❍ Other: ______________________________

4. **Name the faculty/department of all post-secondary programs you have been enrolled in, now or in the past.**

   ____

5. **Were any of these post-secondary programs related to STEM (science, technology, engineering, and math) fields?**

   ❍ Yes

   ❍ No

6. **Is English the first language you learned to speak?**

   ❍ Yes

   ❍ No

   ❍ prefer not to disclose

7. **If not, at what age (in years) did you learn to speak English?**

   ____

8. **Please specify your ethnicity. (i.e. peoples' ethnicity describes their feeling of belonging and attachment to a distinct group of a larger population that shares their ancestry, language or religion).**

   - ❑ White (A person having origins in any of the original peoples of Europe, e.g., German, Irish, English, Italian, Polish, etc.)
   - ❑ Hispanic, Latino, or Spanish origin (A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.)
   - ❑ Black or African American (A person having origins in any of the black racial groups of Africa.)
   - ❑ Aboriginal or Indigenous (A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.)
   - ❑ Asian (A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.)
   - ❑ Middle Eastern or North African (A person having origins in any of the original peoples of the Middle East or North Africa. e.g., Lebanese, Iranian, Egyptian, Syrian, etc.)
   - ❑ Native Hawaiian or other Pacific Islander (A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.)
   - ❑ prefer not to disclose
   - ❑ Other: ____________________

9. **On a scale of 1-7, how interested are you in the topic of rocks?**

   - ❍ 1 - Not interested at all
   - ❍ 2
   - ❍ 3
   - ❍ 4
   - ❍ 5

- ❍ 6
- ❍ 7 - Very interested

10. **On a scale of 1-7, how much do you know about rocks?**
    - ❍ 1 - I know nothing
    - ❍ 2
    - ❍ 3
    - ❍ 4
    - ❍ 5
    - ❍ 6
    - ❍ 7 - I am an expert
11. **How often have you interacted with conversational agents before? (e.g., chatbots, SIRI)**
    - ❍ 1 - Never
    - ❍ 2
    - ❍ 3
    - ❍ 4
    - ❍ 5
    - ❍ 6
    - ❍ 7 - A lot
12. **If you have interacted with conversational agents before, which one(s)?**

    ____
13. **How interested are you in conversational agents?**
    - ❍ 1 - Not at all
    - ❍ 2
    - ❍ 3
    - ❍ 4
    - ❍ 5
    - ❍ 6
    - ❍ 7 - Very interested

B.1.2 Quiz

1. **Which characteristics tell us that a rocks is an Igneous rock? (select all that apply)**

   - ❑ Could have small crystals
   - ❑ Could have large crystals
   - ❑ Could have layers
   - ❑ Could look like glass
   - ❑ Could have holes
   - ❑ Could have sand or pebbles
   - ❑ Could have fossils
   - ❑ Formed from cooling lava and magma
   - ❑ Formed when pieces get deposited and compacted
   - ❑ Formed from heat and pressure

2. **Which characteristics tell us that a rock is a Metamorphic rock? (select all that apply)**

   - ❑ Could have small crystals
   - ❑ Could have large crystals
   - ❑ Could have layers
   - ❑ Could look like glass
   - ❑ Could have holes
   - ❑ Could have sand or pebbles
   - ❑ Could have fossils
   - ❑ Formed from cooling lava and magma
   - ❑ Formed when pieces get deposited and compacted
   - ❑ Formed from heat and pressure

3. **Which characteristics tell us that a rock is a Sedimentary rock? (select all that apply)**

- ❑ Could have small crystals
- ❑ Could have large crystals
- ❑ Could have layers
- ❑ Could look like glass
- ❑ Could have holes
- ❑ Could have sand or pebbles
- ❑ Could have fossils
- ❑ Formed from cooling lava and magma
- ❑ Formed when pieces get deposited and compacted
- ❑ Formed from heat and pressure

4. **Classify the following rocks:**

   (a) Picture of Rock A

   - ❍ Igneous
   - ❍ Metamorphic
   - ❍ Sedimentary

   (b) Picture of Rock B

   - ❍ Igneous
   - ❍ Metamorphic
   - ❍ Sedimentary

   (c) Picture of Rock C

   - ❍ Igneous
   - ❍ Metamorphic
   - ❍ Sedimentary

   (d) Picture of Rock D

   - ❍ Igneous
   - ❍ Metamorphic
   - ❍ Sedimentary

(e) Picture of Rock E

- ❍ Igneous
- ❍ Metamorphic
- ❍ Sedimentary

(f) Picture of Rock F

- ❍ Igneous
- ❍ Metamorphic
- ❍ Sedimentary

5. **[pre-study] Write down as many facts as you know about classifying rocks.**

---

6. **[post-study] Write down as many facts that you have learned today about classifying rocks.**

---

7. **[post-study] Write down as many features as you can remember about shale. [Provide picture of shale used during interaction]**

B.2 Free-form Questions

This section covers the free-form questions used in Chapter 4.

1. **How much did you like teaching Sigma?**

- ❍ Not at all
- ❍ A little
- ❍ Quite a lot
- ❍ Very much

Please explain why/why not:

---

2. **Do you think you were good at teaching Sigma?**

   ❍ Not at all
   ❍ A little
   ❍ Quite a lot
   ❍ Very much

   Please explain why/why not:

   ---

3. **Do you think Sigma was a good student?**

   ❍ Not at all
   ❍ A little
   ❍ Quite a lot
   ❍ Very much

   Please explain why/why not:

   ---

4. **Would you say Sigma had any particular personality or characteristics that stood out to you?**

   ❍ Yes
   ❍ No

   If yes, what is it/are they and how did Sigma express it?

   ---

5. **Is there anything you'd change about the Notebook? Please explain why.**

   ---

6. **Is there anything you'd change about Sigma? Please explain why.**

   ---

7. **Anything else you'd like to tell us about that we haven't asked?**

Appendix C

Study III - Expressive Auditory Gestures

C.1 Questionnaires

This section includes the set of questionnaires used in Chapter 5 that are not otherwise publicly available.

C.1.1 Demographics

1. **What is your gender**

   ❍ man

   ❍ woman

   ❍ non-binary

   ❍ prefer not to disclose

   ❍ prefer to self-describe:

   ____

2. **What is your age?**

   ____

3. **If you are currently enrolled as a student, what degree will you receive upon graduation? If you are not currently a student, what is the highest degree you have completed?**

   ❍ High school graduate, diploma or the equivalent (e.g., GED)

   ❍ Some college credit, no degree

   ❍ Trade/technical/vocational training

   ❍ Associate degree

   ❍ Bachelor's degree

   ❍ Master's degree

   ❍ Professional degree

   ❍ Doctorate degree

   ❍ prefer not to disclose

   ❍ Other: ____________________

4. **Name the faculty/department of all post-secondary programs you have been enrolled in, now or in the past.**

   ____

5. **Were any of these post-secondary programs related to STEM (science, technology, engineering, and math) fields?**

   ❍ Yes

   ❍ No

6. **Is English the first language you learned to speak?**

   ❍ Yes

   ❍ No

   ❍ prefer not to disclose

7. **If not, at what age (in years) did you learn to speak English?**

   ____

8. **Please specify your ethnicity. (i.e. peoples' ethnicity describes their feeling of belonging and attachment to a distinct group of a larger population that shares their ancestry, language or religion).**

   ❑ White (A person having origins in any of the original peoples of Europe, e.g., German, Irish, English, Italian, Polish, etc.)

   ❑ Hispanic, Latino, or Spanish origin (A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.)

   ❑ Black or African American (A person having origins in any of the black racial groups of Africa.)

   ❑ Aboriginal or Indigenous (A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.)

   ❑ Asian (A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.)

   ❑ Middle Eastern or North African (A person having origins in any of the original peoples of the Middle East or North Africa. e.g., Lebanese, Iranian, Egyptian, Syrian, etc.)

   ❑ Native Hawaiian or other Pacific Islander (A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.)

   ❑ prefer not to disclose

   ❑ prefer to self-describe:

   ---

9. **How often have you interacted with conversational agents/voice assistants before? (e.g., chatbots, Apple's SIRI, Amazon Alexa)**

   ❍ 1 - Never

   ❍ 2

   ❍ 3

   ❍ 4

   ❍ 5

❍ 6

❍ 7 - A lot

10. **If you have interacted with conversational agents/voice assistants before, which one(s)?**

    ___

11. **How interested are you in conversational agents/voice assistants?**

    ❍ 1 - Not at all

    ❍ 2

    ❍ 3

    ❍ 4

    ❍ 5

    ❍ 6

    ❍ 7 - Very interested

C.1.2 Quiz

1. **Fossils are found only in**

    ❍ Sedimentary rocks

    ❍ Igneous rocks

    ❍ Metamorphic rocks

2. **Metamorphic rocks are formed by**

    ❍ Extreme heat and pressure

    ❍ Cooling and hardening

    ❍ The breaking down of other rocks

3. **Rocks that are formed from volcanoes are**

   ❍ Gneiss

   ❍ Igneous

   ❍ Sedimentary

4. **Why do vesicles (holes) form in some rocks?**

   ❍ Gas bubbles got trapped in the lava

   ❍ The rock hardened underground

   ❍ Because of pressure on the rock

5. **Lava that cools so quickly that there is little time for crystals to form will lead to igneous rocks that are referred to as:**

   ❍ Porphyritic

   ❍ Intrusive

   ❍ Extrusive

6. **Which rock types can have layers?**

   ❍ Sedimentary & Metamorphic

   ❍ Igneous & Metamorphic

   ❍ Only Sedimentary

7. **What is the texture of this sedimentary rock? (contains shells/fossils) [show rock image]**

   ❍ Biochemical

   ❍ Chemical

   ❍ Clastic

8. **What is the texture of this metamorphic rock? [show rock image]**

   ❍ Nonfoliated

   ❍ Foliated

   ❍ Phaneritic

9. **Put the following in the order in which they can occur: erosion, transport, weathering, compaction, burial, cementation, sedimentary rock.**

   ❍ Erosion, transport, weathering, compaction, deposition, burial, cementation, sedimentary rock

   ❍ Erosion, weathering, transport, deposition, compaction, burial, cementation, sedimentary rock

   ❍ Weathering, erosion, transport, deposition, burial, compaction, cementation, sedimentary rock

10. **What is the texture of this metamorphic rock? [show rock image]**

   ❍ Foliated

   ❍ Nonfoliated

   ❍ Phaneritic

11. **Which type of metamorphism is associated with mountain building?**

   ❍ Contact

   ❍ Burial

   ❍ Regional

12. **This rock cooled on the outside of a volcano and contained air bubbles when it was forming. What type of rock is this?**

   ❍ Extrusive igneous rock

   ❍ Phaneritic igneous rock

   ❍ Biochemical sedimentary rock

13. **Why does weathering have to occur first in the formation of some sedimentary rocks?**

   ❍ Because clastic sedimentary rocks are formed from pieces of other rocks that have been cemented together

   ❍ Because biochemical sedimentary rocks are formed from pieces of other rocks that have been cemented together

   ❍ Because chemical sedimentary rocks are formed from pieces of other rocks that have been cemented together

C.2 Free-form Questions

This section covers the free-form questions used in Chapter 5.

1. **How much did you like teaching Mairi?**

   ❍ Not at all
   ❍ A little
   ❍ Quite a lot
   ❍ Very much

   Please explain why/why not:

   ---

2. **Do you think you were good at teaching Mairi?**

   ❍ Not at all
   ❍ A little
   ❍ Quite a lot
   ❍ Very much

   Please explain why/why not:

   ---

3. **Do you think Mairi was a good student?**

   ❍ Not at all
   ❍ A little
   ❍ Quite a lot
   ❍ Very much

   Please explain why/why not:

   ---

4. **Would you say Mairi had any particular personality or characteristics that stood out to you?**

   ❍ Yes

   ❍ No

   If yes, what is it/are they and how did Mairi express it?

   ---

5. **Would you say Mairi expressed any emotion(s)?**

   ❍ Yes

   ❍ No

   If yes, what emotion(s) and how did Mairi express them?

   ---

6. **Is there anything you'd change about Mairi? Please explain why.**

   ---

7. **Anything else you'd like to tell us about?**
