Towards Automatic Generation of Peer-Targeted Science Talk in a Curiosity-Evoking Virtual Agent

Leave Authors Anonymous
Institute Name
City, Country
emailaddress

Leave Authors Anonymous
Institute Name
City, Country
emailaddress

Leave Authors Anonymous
Institute Name
City, Country
emailaddress

ABSTRACT

Curiosity is a critical skill that spurs learning, but is often found to decline with age and schooling. Recent research has shown that peer interaction may serve a special role in inducing curiosity through increased uncertainty and conceptual conflicts, since peers have similar knowledge and ability. For a virtual agent to stimulate curiosity in peer-interaction contexts , the virtual peer should be able to engage in curiosity-scaffolding behaviors such as argumentation and hypothesis formulation. Consequently, the agent must be capable of on-task and open-ended reasoning required to express these verbal behaviors, while also exhibiting peer-like cognitive abilities. Automatic data-driven generation of curiosity-scaffolding behaviors for virtual peers is important given the spontaneity of child-child conversations, their latent reasoning processes and open-endedness required of curiosity-scaffolding interaction. In this paper, we design and implement a virtual peer that can carry out curiosity-stimulating verbal behaviors while engaged in discussion during multi-party board game play. We use a combination of child knowledge-graph construction and child-child interaction driven modeling to generate game-context appropriate behaviors that are compatible with 9-14 year old children. Encouraging human evaluation of the generated behaviors and generalizeability of the generation framework to other tasks opens up new directions in incorporating open-endedness and science talk in virtual agents that will make them truly play a peer role in learning.

KEYWORDS

Virtual Peer, Curiosity, Open-Ended Play, Board game play, Knowledge Base, Semantic Memory,Behavior Generation, Cognitive Architectures

ACM Reference Format:
Leave Authors Anonymous, Leave Authors Anonymous, and Leave Authors Anonymous. 2018. Towards Automatic Generation of Peer-Targeted Science Talk in a Curiosity-Evoking Virtual Agent. In *Proceedings of Intelligent Virtual Agents (IVA'18).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.475/123_4

IVA'18, August 2018, Sydney , Australia

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

1 INTRODUCTION

Some of you probably sat alone taking apart clocks or radios to see how they worked. Others may have sat with friends or classmates to build a computer from scratch. Following prior work, we define *curiosity* as this strong desire to learn or know more about something or someone [?]. Curiosity often develops in response to external stimuli that evoke knowledge gap and knowledge dissonance [23]. It is one of the important social-emotional learning skill that leads to exploratory knowledge seeking [1]. Recent studies show that interpersonal knowledge seeking behaviors exert a stronger influence on curiosity than oneś own behaviors, during small group learning among 5th and 6th graders[48]. In particular, behaviors that elicit uncertainty and conceptual conflict amongst peers, such as question asking, argumentation, hypothesis verbalization and justification, tend to evoke not only individual, but joint curiosity across group members[36].

Game play provides special opportunities to assess and support curiosity through a safe and playful environment to experience uncertainty and exploration, in the individual[21] and group[52] setting. The scaffolding for "social connectedness and meaningful participation" that a game provides[18] also makes it an engaging activity for virtual peers that simulate the appearance and ability of a real child. Peer scaffolding has been shown to support positive development in children for curiosity[15], growth of mind[37], social interaction[3], and literacy[44].

The open-ended nature of curiosity raises two challenges for a virtual peer to fulfill curiosity-stimulating social scaffoldings. First, the virtual peer has to be capable of engaging in behaviors that provide open-ended possibilities to evoke knowledge gap and uncertainty[30], such as creating new hypotheses or arguing for alternate viewpoints. Open-ended possibilities create opportunities for others to respond to uncertainty and alternatives and in turn exhibit similar behaviors. Second, the virtual peer has to demonstrate equal abilities, in order to elicit the kind of cognitive dissonance that evokes curiosity, since children tend to challenge and compare the correctness of one another's ideas, but may accept adults' ideas unthinkingly due to their high knowledge authority[38, 41].

In this work, we study the virtual peer's ability to scaffold curiosity in the context of a custom-designed cooperative board-game. The game incorporates uncertainty by asking players to explore a fictional space through question asking and discussion, and then to take calculated risks based on what they think they know. In the game, both game state updates (e.g. drawing cards, using resources) and curiosity-related player behavior (e.g. generating hypotheses, asking questions, making arguments) serve as game moves. There is therefore a need for real-time reasoning that processes both spontaneous peer interactions and dynamic game state updates; A limited repertoire of scripted conversation moves or Wizard-of-Oz manipulation would limit the agent's ability to produce open-ended, peer-like and spontaneous-sounding verbal behaviors during discussion.

Data and machine learning driven generation of verbal behaviors through reasoning over social interactions and current game state is a more adaptive solution for the spontaneous generation of behaviors with open-ended possibilities and peer-like characteristics. In the current work, we develop a fully data-driven technique for the generation of key curiosity-inducing behaviors. We generate verbal behaviors to be game-context related and peer-targeted. More specifically, we construct a distributional semantic knowledge representation for the target age-group, that the virtual agent can query based on the contextual game information and intended curiosity-evoking behavior. To circumvent the lack of sufficient child-child game-play data, we leverage the use of large, public domain multimedia content available to children for knowledge construction. This affords the agent the capability to generate diverse responses such as new hypotheses and alternatives not seen in previous interactions. To incorporate game-specific constraints that control the extent of open-endedness, we then adapt the generic knowledge representation for the game using game-related syntagmatic relations extracted from game-play data. Finally, in order to generate curiosity related behaviors, we identify important semantic and syntactic patterns that are triggered when these behaviors are displayed by children during game-play. These patterns inform specific search heuristics over the constructed knowledge representation to generate context appropriate behaviors. We also extract language motifs from game-play data as templates to generate the actual natural language utterances for the peer virtual player.

For every step in this generation process that involves data manipulation , we use human evaluation to measure performance of data-driven techniques, thus ensuring that game and context appropriate behaviors are generated despite noisy data and algorithmic imperfections. We discuss the generalizeability of the child-centric adaptations made to virtual agent cognitive architectures and the behavior generation framework that makes game-specific adaptations to general knowledge representation to other collaborative tasks that require open-ended and creative peer interactions.

The main contributions of this paper are two-fold:

(1) We design and implement a novel data-driven approach for the generation of curiosity evoking verbal behaviors and utterances for a virtual peer engaged in a discussion based board game.
(2) We propose a peer-centered design approach for knowledge representation and language modeling that uses a combination on of-task child-child interaction data and child-centered media corpus, and discuss the implications of this approach.

2 RELATED WORK

Peers are individuals of similar age, ability, knowledge, experience and social status [17, 40]. Equal relationship among child peers is particularly beneficial for learning because, when a conceptual conflict or alternative opinion arises, both perspectives have an equal potential to validity without special authority, which may lead to active discussion and argument in resolving and integrating different views [41]. While symmetric age and social status have been generally applied in virtual peers, most of them hold asymmetric knowledge and ability compared to the real child, such as tutors [3], teachable tutees[5, 28], and supportive companions [15, 37]. There are only a few exceptions where the virtual peer is intended to closely simulate child ability. Sam is a virtual peer who can engage in collaborative storytelling and is controlled by young children with autism[51] and Alex is a culturally-authentic virtual peer who engages a child in science talk while switching dialects[7]. Both virtual peers simulate peer-like behaviors but have limited ability to engage in behaviors with open-ended possibilities as their verbal behaviors are pre-scripted or extracted from the limited child-child interaction corpus.

Cognitive architectures provide the core frameworks behind the reasoning modules in virtual agents. Most cognitive architectures focus on simulating optimal intelligence. There are only a few studies that investigate cognitive architecture through the lens of development and individual differences[22, 31]. They identify three main cognitive factors that influence human intelligence - knowledge change, memory and processing capacity, and strategy choice with experience. While there is a lack of precise understanding of the number of knowledge structures that can be activated and processed at one time that match with different developmental stages, media exposure is considered one of the main sources for childrenś knowledge acquisition [46], and data-driven behavior modeling has provided a way to simulate non-optimal cognitive processes of strategy choice and language generation in scenarios such as tutor-tutee interaction[34], and previously mentioned child-child interaction in science discovery and storytelling[7, 51]. We propose a fully data-driven adaptation of the cognitive architecture model towards the generation of peer-like behaviors with open-ended possibilities.

Semantic memory stores the required reasoning module of the virtual agent for generation of these complex behaviors from the current game context. The virtual agent's semantic memory encodes word meanings, facts, concepts, and general world knowledge required for day-to-day reasoning. Semantic memory has traditionally been represented as a semantic graph consisting of nodes (concepts and terms) and edges(relationships)[27]. For example, the concepts *dog, animal* are connected by the edge *is-a*. There are several manually curated or automatically generated knowledge bases such as WordNet[13], ConceptNet[29] and NELL[35] etc., that are sometimes directly incorporated into an existing virtual agent dialogue manager [44] for reasoning and next utterance formulation. However, these knowledge bases contain factual information and common sense syntagmatic associations (such as *lions roar*) that are very generic for game play and simulate an adult's semantic memory. Distributional models of semantic memory[**?**] use text corpus-based co-occurrence models like LSA[26] or the more recent Word2Vec[33] to create a semantic vector space in which different concepts(terms) in a large text corpus reside, like **??**. In order to create knowledge graphs that support peer-like reasoning, we draw inspiration from such distributional models. Such techniques have previously been used by [6, 20, 54] to project terms into relation-specific vector spaces using labeled data and supervised machine learning. Our semantic knowledge graph is similarly constructed from a child-centered corpus, and we do so in a semi-supervised manner.

Reasoning over the knowledge graph involves exploring potential and reasonable relations among various entities to respond to a query or complete a task. Open-endedness of discussion based dialogue means that different kinds of relationships can emerge. The game we consider, however, provides a context within which relatively constrained discussion takes place, thereby limiting the type of relationships to reason over. We again rely on data to learn the relationships deemed important for game play and for curiosity-stimulating behavior generation. Automated biological hypothesis generation *OpenCog* also utilizes relationships that are extracted automatically from free text[14]. Most automatic relation extraction methods like these use dependency parsing to break sentences down to their syntactic tree-structures to extract underlying relations as shown in ??. For instance,[11] design a semi-supervised method for extracting protein-protein relations based again on dependency parse trees. The Episodic-logic framework[45] uses alternative parsing methods based on propositional logic that create a semantic representation of natural language. Inspired by these works, we develop a generalized approach based on dependency parsing for behavior-specific relation extraction.

The generation of natural language utterances requires the construction of a child-centered language model to capture common lexical patterns children use during game play. Due to a lack of sufficient child-child data for training an automatic natural language generation model, statistical pattern-mining based approaches are a more feasible strategy. [4] extract response structures from previously annotated sequences to generate factoid questions. [10] investigate and extract dialogue patterns from human-human interactions to be used by a software agent to interact with real humans . [42] construct sentence patterns in the form of sequences of parts-of-speech and a simple lexicon of words and then populate template sequences with appropriate words from the lexicon. In a similar fashion, we design our natural language generation method through extracting lexical patterns and templates in a data-driven manner, which are then appropriately completed with content words based on the agent's reasoning.

3 METHOD

In this study, we build a virtual peer that can generate key curiosity-inducing behaviors including justification, argumentation and hypothesis verbalization to elicit uncertainty and conceptual conflict in group members. We used a child-centered modeling approach to enable the virtual peer to think and behave like a competent 9-14 year old child in a conversation-based collaborative game called Outbreak [52]. Outbreak is a question-asking and discussion driven board game for a group of two to five players to collaboratively investigate hidden threats in a series of broken science labs (threats such as haunted by a ghost, leaking chemicals etc.). Each time the players enter a new room, they ask questions using provided question templates to the game master (e.g. What happens if..., Is the room...) for up to two minutes. After the question asking phase, players enter the discussion phase when they have to collaboratively decide the right resource cards with skills needed to conquer the threats. The 7 skills in Outbreak are *fight, love animal, hack*

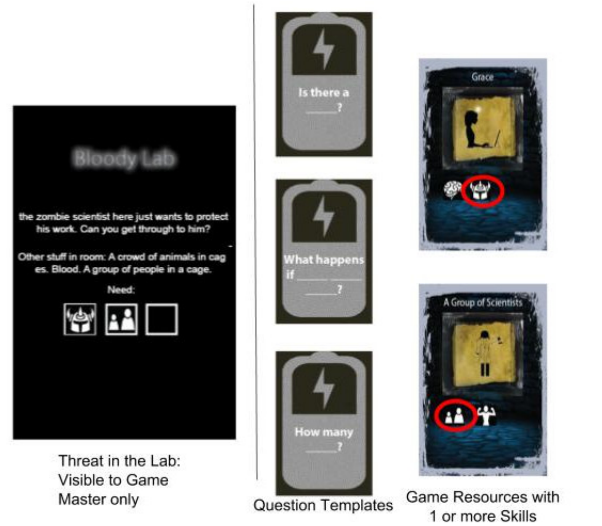

Figure 1: Outbreak Game Play Scenario (required skills circled)

(computers, software), block, run fast, friend and see. We chose Outbreak as the study activity because it provides an engaging and exploratory experience with sustained level of uncertainty, and allows the virtual peer to carry out key curiosity-inducing social scaffoldings to validate curiosity elicitation strategies. An example game scenario is provided in 1 and a small conversational snippet is illustrated below:

Question asking phase:
Player 1: "Is there a zombie in the room?"
Game master: "Yes, there is a zombie in the room."
Discussion phase:
Player 2: "I think we should use the helmet because she said there is some sort of zombie in the room, it might eat our brains, (laughter)."

Outbreak Data Collection. We collected child-child interaction data of 10 groups of 3-4 players, 9-14 years old (30 participants in total with 13 female participants) playing Outbreak in a controlled lab setup. We recruited from CITY NAME public and charter schools, and a YMCA community center in a historically under-resourced neighborhood. All participants' parents gave consent, and participants gave minor assent. The confederate experimenter first runs a scripted practice round to explain the game rules while also playing as the game master. Participants then play the Outbreak game for either 40 minutes or until they complete all rooms and reach the end of the game. Participants then complete a self-report questionnaire for affective arousal. We used four camcorder recorders, four webcam devices and a fisheye camera to record the video data including the front face and group view of each participant and a top-down view of the table and game board. The audio data is recorded using lapel microphone attached to the collar of each participant. We transcribed and annotated a convenience sample of the first six groups of the game play.

Verbal Behavior	Definition
Justification	Showing something to be right or reasonable by making it clear
Argument	A coherent series of reasons or facts to support or establish a point of view
Hypothesis Verbalization	Expressing one or more different possibilities or theories to explain a phenomenon by relating two variables

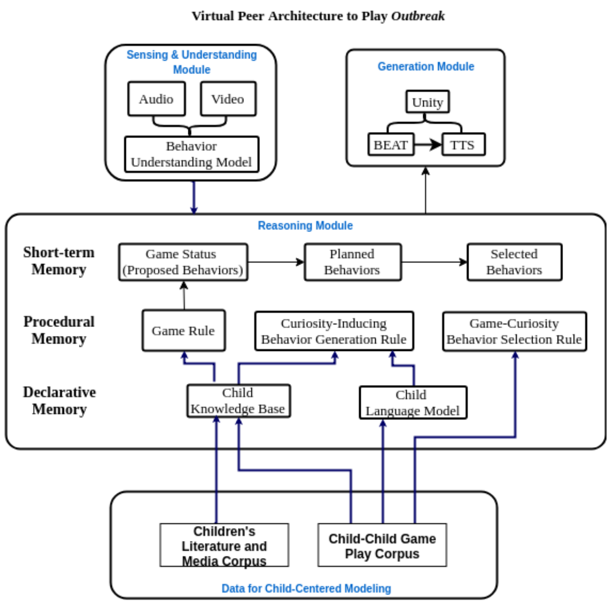

Figure 2: System Architecture of the Virtual Peer

We used human annotators to annotate every clause in our corpus for three verbal behaviors that have been correlated to increased curiosity in past work[47, 48] - Justification, Argumentation and Hypothesis Verbalization. Brief definitions for these behaviors are provided in Table **??**. Inter rater reliability (Krippendorf's alpha) for each of these annotations was above 0.7. Additionally, we also annotate for game resources such as question templates, skill cards and possible threats (termed as keywords) that players refer to in these utterances.

We describes the cognitive architecture of the virtual peer in 2, that refers to key components of general cognitive architecture proposed in ACT-R[2], Soar [25] and CLARION[50] along with a tier for the data-driven adaptation. The cognitive architecture includes the following modules (1) sensing - updates game status by tracking the game elements on the table using marker-based computer vision technology; (2) reasoning - plans and selects game appropriate verbal behaviors that provides open-ended possibilities and eliciting curiosity; (3) behavior generation - realize the verbal behaviors through text-to-speech (TTS) using Amazon Polly, and associated non-verbal behavior generation using Behavior Expression Animation Toolkit (BEAT) [8] and Unity game engine. In this paper, we focus on describing the reasoning module that enables the generation of required game behaviors, namely question asking using question templates and suggesting a skill or card during the discussion. We additionally enable reasoning for and generation of curiosity-inducing verbal behaviors such as arguments and hypothesis verbalization. We explain pertinent parts of the reasoning module in detail.

Semantic and Procedural Memory. Short-term memory consists of the agent's current belief state for the game, the next chosen behavior and utterance. This memory is constantly updated with external stimuli and is refreshed at the beginning of every new game round. Semantic memory encodes word meanings, facts, concepts, and general world knowledge that agent uses to reason about the game, while procedural memory maintain rules that are required for planning the content and natural language utterance of different verbal behaviors.

Processing Unit. The processing unit refers to the key reasoning steps over the components of the cognitive architecture involving a) elaboration - monitoring short-term memory, b) operator proposal - propose operators appropriate to the current situation, and c) operator selection - select the optimal operator. The corresponding processing unit in our cognitive architecture includes the following steps:

(1) **Game Belief Updating** - updates the virtual peer's belief of the game status; For example, Player 1 asks the game master: "Is there a robot?" and the game master responds: "Yes". The game belief updating module will register the question and answer pair, and inquire the game rules and child knowledge base to update the possible set of game resources/skills that associate with the keyword "robot" (e.g. fight, be friend, block)
(2) **Behavior Content Planning** - Plans for the content of legitimate curiosity-inducing verbal behaviors according to the game status updates. For example, in the discussion phase, because the game master provided the information that there is a robot in the room, Player 1 suggests: "I think we should use fight because she said there is a robot". The virtual peer can provide alternative opinion through argument: "wait, we will need hack because the robot could be broken", or hypothesis generation: "what if the robot is not dangerous? It may be lonely"
(3) **Behavior Selection** - Selects appropriate verbal behaviors to generate that fulfill the purpose of game play with an emphasis on curiosity-induction. For example, statistically, the child may be more likely to make arguments following another person's suggestion than hypotheses, so the planned argument behavior is chosen for generation more often than hypothesis verbalization.

The game belief update and behavior planning are driven by a child knowledge graph, a distributional representation of semantic memory that is extracted from open domain child-centered media data and adapted for the Outbreak task into an associative mapping between skills and different keywords.

3.1 Encoding Children's Semantic Knowledge

As previously described, the semantic memory (referred to as the knowledge base in our framework) connects entities using relations

Table 1: The Detailed Information of Collected Dataset

Data Source	Size
Kindle Ebooks	9.67MB
Online Free Books	46.4MB
Board Game Rule Books	778KB
YouTube Videos Scripts	8.76MB
Facebook Children's Literature Dataset	26.6MB
Total	**92.3MB**

that represent a shared encoding of their meaning. Meaning lies at the cornerstone of nearly all aspects of human cognition and a knowledge graph encodes the relative meaning of concepts accrued from repeated episodic experiences. Hence, we begin with the construction of a general child-centered knowledge graph and develop strategies to traverse this graph to respond to different situations and generate corresponding behaviors.

Collecting Auxiliary Dataset. Our existing child-child game-play data consisting of real children's conversations can be prohibitive to the construction of a knowledge graph due to a) small size, and b) children using repeated skills and keywords across different game sessions. We augment this dataset with larger and more generic textual data and ensure that this data contain the knowledge that real children of the target age group are exposed to[46], so as to be able to encode conceptual associations that a human peer is likely to make. We reviewed articles and surveys about children's media usage patterns and preferences for specific genres. For instance, [39] find that children in the age group of 10-18 prefer FM Radio and Television as their favorite digital media. [43] find that about 84% of children in their study had access to home Internet and 81% of them watched videos on the Internet, indicating that online videos have become more popular among the youth as a source of information. Besides, e-books also attract a substantial number of children and are gradually replacing traditional paper-backs. Along the dimension of genre, we explored the kinds of information that children are most interested in. [49] conducted a survey of reading preferences for children aged 2-18. Their results show that the top 3 categories that children are most interested in are animals, science, and sports. Davila et al. review the authorized reading materials and actual reading materials preferred by children in their work [9]. Their study reveals that boys and girls equally like fictions which contain horror, humor and adventure. They also suggest that the list of bestselling children's books is closely related to young audiences' preference, which inspired us to refer to the bestseller ranks of books in recent years. Based on the above review of research and surveys, we collected text data covering popular media types (e-books, video transcripts and board game rule books) and genres (like fantasy, history, science and so on). We also included the children's literature corpus released by Facebook consisting of fairy tales and story books from Project Gutenberg[16]. Amount of data of each media type is listed in 1

3.1.1 Vector Spaces as Child Knowledge Representation. We train a distributed word embedding model (Word2Vec[33]) on the collected children's corpus to transform words into vector representations. In distributed models, the semantics or meaning of the word is distributed across all dimensions of the vector that encodes it. In the resulting vector space, plenty of linguistic regularities and patterns are encoded through the learned word vectors. The biggest merit of such models is that semantically similar words are located close to each other and tend to form clusters. This characteristic is useful for exploring and discovering new pairs of words that share similar semantic relations. We exploit this characteristic to adapt this generic semantic vector space to the Outbreak task.

3.1.2 Task-Specific Transformation of Knowledge Representation.

A Keyword - Skill Mapping for Outbreak. A competent player of the Outbreak game should be able to choose and suggest a reasonable skill card given the current game context in order to overcome potential threats. The game context information consists of threats (keywords) mentioned by the game master and other players in previous turns and provides clues to the virtual child to select appropriate skill cards. For example, if someone asked if there is a *computer* in the room, here *computer* is a keyword, and the virtual child may suggest the skill *hack* as a reasonable response since it relates to *computer*. The knowledge graph constructed previously can be too generic to relate these two concepts since they may not have co-occurred in the larger child-centered corpus. Hence, we use a semi-supervised heuristic to map skill words and plausible keywords in the generic knowledge graph, denoted by K. We use a small set of known mappings between skills and keywords extracted from the Outbreak game play corpus as a seed. Our heuristic method can be represented in the following formulation:

$$argmin[\alpha(\parallel x - x_c \parallel) + \beta \parallel x - x_s \parallel] \qquad (1)$$

, where x is any vector in K, x_c is the vector centroid of keywords that were mapped to a given skill s, and x_s is the word vector of the skill word. The first part of this formulation finds the closest neighbors of the keywords that were linked to a particular skill since we hypothesize that such words are also related to the skill. The other part selects terms that are closest to the corresponding skill. α and β control the relative strength of association of a new potential keyword to already existing keywords and the skill itself. The method is also illustrated in Figure 3. Table 2 shows top retrieved potential keywords for a few skills. This process results in links between potential new keywords and skills being established apart from those already present in the limited game play data. This is useful for simple verbal behaviors like making a suggestion to put forth a skill when a new keyword is encountered during future games. Our virtual agent would suggest the skill that is most strongly mapped to that keyword by 1, so the virtual agent may say "it would be smart to put love animals because there is a dog" given the keyword *dog*. During the Q&A phase of the game, the question templates can be completed by picking a new related keyword for a certain skill and filling it into the template. For example, if the virtual agent has the skill *see* and a question like "Does the room look dark?" might be asked. These keyword-skill mappings enable the virtual agent to engage in basic game play with real children. We evaluate these mappings manually, which we describe in Section 4

Table 2: The Top Keywords For a Part of Skill Words

Skill	Fight	Love Animals	Hack	See
Keywords	Enemy	Animal	Robot	Escape
	Attack	Dog	Machine	Light
	Terrible	Creature	Computer	Dark
	Deadly	Cat	Software	Get away
	Monster	Bird	Automatic	Bright
	Powerless	Beast	Setup	Moonlight
	Danger	Kitten	Engine	Sunlight
	Desperate	Monkey	Mechanical	Run
	Kill	Snake	Radio	Dim
	Destroy	Alive	Wires	Sight

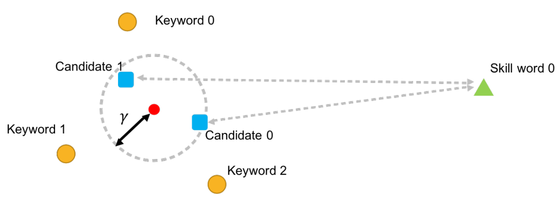

Figure 3: The Figure of Proposed Method.

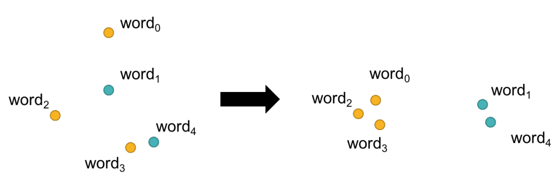

Figure 4: Retrofitting Knowledge Graph to be more game-specific: The dots with the same color are mapped to the same skill

Adapting Knowledge Base for Outbreak. During the course of game play, we expect certain associations in the knowledge graph to be more reinforced than others and a gradual transformation of the knowledge base to optimize strategies for Outbreak that result in more wins. This observation is drawn from the memory activation hypothesis wherein the strength of association between certain cognitive units increases with practice and repeated tasks, forming a *working memory* for the task[**?**]. For instance, children often start associating words like "Zombie" and "Monster" with "fight" or "block" as they play more rounds. We adapt the general knowledge network to a game-specific one using the retrofitting method from [12] to modify the vector space based on the newly extracted and evaluated associations. The idea of retrofitting word vectors is to force a word to be close, not only to its original neighborhood, but also to other concepts that share *special* edges with it. We pictorially depict this transformation in Figure 3.

3.2 Curiosity-related Verbal Behaviors Generation

Our ultimate goal is for the virtual child to induce curiosity in other players during game play, and previous work [36, 48] has shown that three verbal behaviors in particular - Justification, Argumentation and Hypothesis Verbalization - can stimulate curiosity during peer-peer interaction. While the generated keyword - skill associations can already produce simple verbal behaviors to support basic game play, we still hope to support higher order reasoning that is crucial to elicit curiosity stimulating behaviors. In this section, we describe how our pre-built knowledge network can be leveraged to generate these three verbal behaviors.

3.2.1 Extracting Dependency Relations for Behavior Generation. Inspired from related work in the area of automatic relation extraction, we begin by doing a dependency relation based syntactic analysis of the game-play data to discover common syntactic structures children use when they display these three verbal behaviors. However, the size of our child-child data is limited and curiosity-stimulating behaviors are sparse, which makes the analysis results less convincing. Once again, we turned to auxiliary datasets that have been created for different verbal behaviors. For example, AI2 Elementary Science Questions corpus was manually annotated for 6 kinds of justification [19] and the Internet Argument Corpus has been annotated for 3079 argument instances [53]. Dependency parsing was applied to both datasets and the frequency of every dependency relation was measured. We find that the dependency relations of *amod* and *dboj* rank in the top 7 among 40 relations. The other top relation types such as *det* and *root* are essential grammatical constructs present in most sentences but do not play major semantic roles. We interpret the high frequency of *amod* and *dobj* as proposing properties of a keyword *amod* or actions that can be done by and to keywords*dobj*. For instance, *a* ***dangerous*** *monster* or *the ghost can* ***kill*** *us*. For hypothesis verbalization, to the best of our knowledge, there is no publicly available auxiliary corpus. Moreover, hypothesis verbalization is rarely done in the outbreak corpus but is found to be specifically associated with making more than one group member curious [36]. [24] claim that a hypothesis is a conjectural statement that encodes the relation between two or more variables, so *amod* and *dobj* can still serve as potential relations between concepts and are useful for generating conjectural statements. For each keyword, we extract various potential relational words based on the above three relations (e.g. for keyword *monster*, an extracted relation word is *dangerous* through the relation *amod*). These relations between words not only serve to discern attributes of objects or actions they can do, but may also be the latent reasoning that supports the choice of a specific skill card.

3.2.2 Curiosity-related Verbal Behaviors Generation. The general strategy for behavior generation involves using the current game context (keywords mentioned thus far during game play) to construct queries and search the game-specific child knowledge graph. The most generalizable search strategy we develop is the bottom-up strategy: Given a keyword and skill, we find terms related by the *amod,dobj* relations that are semantically similar to both, the skill and the keyword. For search, we start from the keyword, and iterating over top semantically similar relation words,

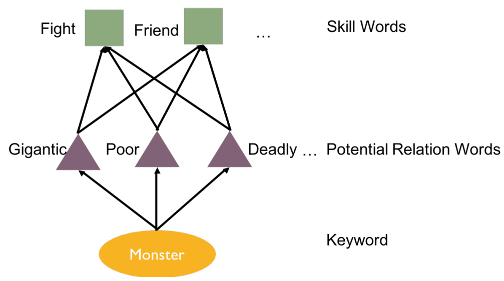

Figure 5: The Figure of Bottom Up Strategy in an Example

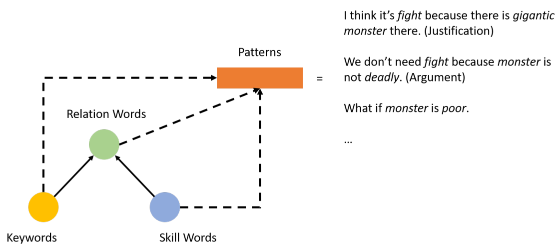

Figure 6: The Figure Showing the Strategy for NLG

we calculate their semantical similarities with the skill using the following formula.

$$\alpha * sim(relatiion, skill) + \beta * sim(keyword, skill) \qquad (2)$$

,where the sim(x, y) calculates the cosine similarity of vectors for words x and y in the retrofitted task-adapted word vector space, and α and β are used for controlling the relative strength of each association. Figure 5 shows an example of the bottom up strategy and the combination of the strategy with knowledge graph for final utterance generation is shown in Figure 6.

For **Justification** we pick a keyword in the known game context and use the knowledge graph to find an appropriate skill as before. An acceptable reason is generated to support this associated using the bottom-up approach. For example, if *monster* is a mentioned keyword and *fight* is chosen as a possible skill to overcome the *monster*. Words related to *monster* and *fight* are scored and ranked using 2 and a top-scoring term is chosen to complete the justification. A possible generated sentence can be: "I think we need *fight* because there is *gigantic monster* there".

Argument occur when one gives reasons to disagrees with someone else's ideas. We find from child-child game play data that children often argue against a skill suggested by other players. Suppose we want to argue against a previous justification. So in this case we are supposed to focus on the skill word *fight*, and based on the mentioned keyword, say *monster*, we calculate the scores for relation words as done in justification. After filling these words into an argument pattern, the result may be like: We don't need *fight* because *monster* is not *deadly*. Providing an alternative perspective rather than negative evaluation can also be another strategy for argumentation. For this case, we start from the keyword but then choose the next closest skill other than the one mentioned in the context. For example, the next relevant skill word for *monster* is *friend*, and the final sentence is likely to be: No, what if the *monster* is *wretched*?

Table 3: Examples of Generated Patterns

Verbal Behaviors	Generated Patterns
Justification	*KEYWORD* is *RELATION* so we need *SKILL*. I put *SKILL* down because it has *RELATION KEYWORD*.
Argument	There's no *RELATION KEYWORD* there. No *SKILL*, because this has no *RELATION KEYWORD*.
Hypothesis Verbalization	Try *SKILL* what if there's *RELATION KEYWORD* there. Maybe like *RELATION KEYWORD*.

Hypothesis verbalization always refers to expressing some unseen potential possibilities by providing a relation between two entities. Using the above strategy, we can propose an unmentioned angle or perspective about one keyword as a hypothesis. For instance, the keyword *monster* is usually related with *fight*, but there is another alternative that we select *friend* with the strategy to generate a hypothesis verbalization like: But what if *monster* is *poor*.

Apart from the bottom-up strategy, we also think of a top-down strategy for creating more diverse sentences for verbal behavior generation. This idea comes from one important observation about the Outbreak data viz. that sometimes children may not use information they acquire during the Q&A phase to relate entities and use a more direct approach for their reasoning, which starts from the skill cards they hold in their hands and try to form some plausible reasons that allows them to use these cards. Here, for each skill word, we calculate the scores of a combination of relation words and keywords closely associated with the skill, and rank the resulting list. This strategy is especially helpful for hypothesis verbalization and question asking, which can be generated based on initiation from the skill words with little past game context information. For instance, for the virtual child to use the hack skill, our method may output a sentence like: *There might be an automatic door in the room.* Such a sentence is a logical hypothesis about what may exist in the room and promotes other children to think more about other plausible threats. A detailed example is shown as Figure 8.

Natural Language Utterance generation. We also try to create sentence patterns for the generation of final surface forms(agent utterance) of these verbal behaviors. Another dependency parsing was performed on Outbreak game data, and for each verbal behavior, we pick the Top 10 common syntactic structures and extract the corresponding natural language instances that contain these syntactic structures. These patterns are made generic by replacing the key content words with placeholders like *SKILL* and *KEYWORD*, to create templates that can be filled in real-time to generate new utterances.

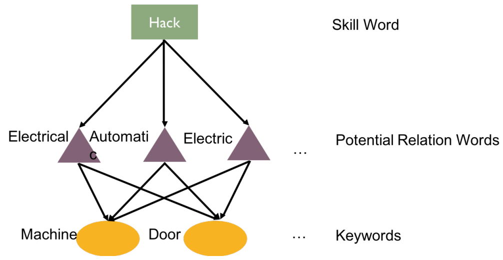

Figure 7: The Figure of Top Down Strategy in an Example

Table 4: Examples of Real Children's Dialogues and Generated Sentences

Verbal Behaviors	Real Sentences	Generated Sentences
Justification	*Something* is *broken* so we need *fix.*	I think it's *fight* because there is *gigantic monster* there.
Argument	We won't need it because there's no *strength.*	We don't need *fight* because *monster* is not *deadly.*
Hypothesis Verbalization	Maybe a *computer*'s *falling.*	What if *monster* is *poor.*

4 EVALUATION

Game Belief Update Evaluation. Precision and Recall are adopted to evaluate the newly constructed keyword-skill associations. For precision, we generated top 20 potential keywords for the 7 skills in Outbreak and asked 4 in-house annotators to recognize reasonable pairs based on their common-sense judgment of a reasonable semantic relationship between a keyword and skill in the context of the game. Since the definition of reasonable can be complex and subjective, we disambiguate using majority vote among human annotators and leave out cases where even human annotators disagree. Thus, we calculated precision as the proportion of confident reasonable mappings that the computational model is able to correctly retrieve and get a score of 0.8344. The Krippendorff's alpha for annotators' inter-rater reliability is 0.8636. For Recall, we randomly picked 200 words from the larger peer-targeted corpus and asked annotators to map each word of ne or more of the 7 skills. We again use majority vote as the ground truth for reasonable mappings. The Krippendorff's alpha was 0.8600. Our computational model is used to find the corresponding highest ranked skills for these keywords and to measure how many of one or more of them are correctly retrieved. The recall was 0.6316.

Curiosity-related Verbal Behaviors NLG Evaluation. We evaluated the generated curiosity-related verbal behaviors for game and context appropriateness. We extract artificial game contexts in the form of keywords mentioned by other players/game master as the potential context to reason over and plan behavior content from. We generated 100 sentences for three verbal behaviors and also mixed them with 100 randomly picked sentences from child-child game play corpus. We asked the annotators to classify the behaviors as Justifications, Hypothesis Verbalization and Argumentation based on the generated utterance and the available context. The computed accuracy for classification is 0.76 and the Krippendorff's alpha for agreement among annotators is 0.6536.

5 DISCUSSION

Curiosity-Inducing Behavior Generation. We validate that the generated skill-keyword associations and verbal behaviors are relevant to the current game context. The evaluation performance of the skill-keyword association measures the success of automatically extracting associative relationships between concepts in the game using a game-adapted child-centered textual dataset. Even though skills and key-words are specific to the Outbreak game, we present a generalizable method of building a semantic memory representation (knowledge graph) from a large target-appropriate text corpus that uses an initial seed of syntagmatic associations from human-human on-task data. The knowledge graph serves as the primary source for task-based reasoning. This technique is generally applicable to a wide range of tasks that require the virtual agent or dialogue agent to reason over structured knowledge in a constrained context. Incorporation of structured knowledge into virtual agents is an emerging field for addressing the challenges of personalization, intent understanding from context and semantic relevance of responses[**?**]. Recently, agents are using distributional semantic memory for small reasoning tasks - robots that understand analogies in human instructions[**?**] or agents that detect behavioral affordances such as objects that can be grasped, drunk, worn, etc[**?**]. Children also acquire semantic knowledge from visual input and a potential future direction is to encode textual and visual information in a common semantic space (as done in [**?**]) for the virtual agent to reason over multi-modal context.

We automate the agent's curiosity-related verbal behavior generation for a limited number of syntactic relations between entities. The promising evaluation of the generated curiosity-related behaviors for game-context appropriateness can potentially support the automated generation of such behaviors in constrained task settings. For instance, we can generate arguments, hypotheses and justifications for the agent based on keywords extracted from conversational history. These behaviors are integral to scientific talk and can be incorporated into intelligent pedagogical agents that engage in educational games or peer-tutoring. Traditional knowledge bases support basic semantic relationship such as *type of, synonyms, antonyms etc* while our method supports syntactic relations to generate context relevant and coherent sentences. Understanding causal relations between concepts and the pragmatics of human conversation are still unsolved, but crucial for engaging humans in scientifically accurate talk.

In the adapted cognitive architecture proposed in this work, we have developed the module for generation of curiosity related behaviors. The procedural memory we propose also includes a model that selects the next behavior that can fulfill the purpose of game play or stimulate curiosity in other players. In future work, we plan to build a behavior selection model that optimizes for positive change in group curiosity (curiosity of all group members) to carry out a full scenario testing of Outbreak game play with real children. This will be done in order to validate if curiosity-stimulating behaviors selected and generated by the virtual peer can stimulate curiosity in real children.

Peer-Targeted Knowledge Construction and Language Modeling. Augmentation of child-child interaction data with child-centered media data allows for age-appropriate associations to be learnt by our model. This is done to ensure co-equal participation of the virtual peer to facilitate constructive debate of ideas instead of acceptance from an agent with higher authority. This is not only crucial for effecting positive impact on curiosity, but also leads towards the general modeling of virtual peers that display symmetric age-appropriate cognitive abilities. We currently evaluate our approach for age-appropriate associations and peer-like behaviors and linguistic patterns based on adults' perceptions of children's reasoning and language. A thorough evaluation of the relative success of this data-driven technique should involve children of the target age in the evaluation loop. Children may give implicit evaluations for believability and age-appropriateness during their interactions with the virtual peer. Crowd-sourcing with child workers has just recently been explored [32] in the CHI community. Manojlovic et. al have found that joint tasks assigned to parents and children are more acceptable. There is potential new space to generate creative tasks for children to evaluate the performance of data-driven techniques for modeling of virtual peers that can also serve as additional means to collect target specific data. This can further fine-grained incremental development of virtual peer technology rather than final user studies with children.

6 CONCLUSION

Recent research has shown that peer interaction may serve as special stimuli in inducing curiosity by facilitating increased uncertainty and conceptual conflicts. Collaborative games provide a safe and fun setting to express uncertainty and impose the need for social connectedness and participation that virtual agents can leverage. Endeavoring towards this goal, we build a virtual peer agent that can elicit curiosity scaffolding behaviors while engaging in an open-ended discussion based board game. We develop and implement a behavior generation module to realize age-appropriate curiosity-inducing behaviors such as question asking, hypothesis verbalization, argumentation, and justification during open-ended discussion in a multi-party board game. We use a combination of child knowledge base induction, game-specific adaptations to the knowledge base and child-child corpus driven behavior and language modeling to generate spontaneous and diverse curiosity-inducing behaviors. Promising human evaluation of different stages of the generation process for game-context appropriateness and generalizeability of the generation framework to other tasks and age-groups opens up encouraging new directions in peer virtual agent modeling for open-ended game play and science reasoning tasks. We present the challenges involved in measuring age-appropriateness and open-endedness of generated conversations and propose to implement behavior selection to carry out user testing for extrinsic validation of curiosity-related behavior generation.

REFERENCES

[1] 2016. World Economic Forum (2016) New Vision for Education: Fostering Social and Emotional Learning through Technology. http://www3.weforum.org/docs/WEF_New_Vision_for_Education.pdf

[2] John R Anderson, Daniel Bothell, Michael D Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. 2004. An integrated theory of the mind. *Psychological review* 111, 4 (2004), 1036.

[3] Sara Bernardini, Kaska Porayska-Pomsta, Tim J Smith, and Katerina Avramides. 2012. Building autonomous social partners for autistic children. In *International Conference on Intelligent Virtual Agents*. Springer, 46–52.

[4] Essia Bessaies, Slim Mesfar, and Henda Ben Ghezala. 2017. Generating Answering Patterns from Factoid Arabic Questions. *Proceedings of the Linguistic Resources for Automatic Natural Language Generation-LiRA@ NLG* (2017), 17–24.

[5] Gautam Biswas, Krittaya Leelawong, Daniel Schwartz, Nancy Vye, and The Teachable Agents Group at Vanderbilt. 2005. Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence* 19, 3-4 (2005), 363–392.

[6] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*. 2787–2795.

[7] Justine Cassell, Kathleen Geraghty, Berto Gonzalez, and John Borland. 2009. Modeling culturally authentic style shifting with virtual peers. In *Proceedings of the 2009 international conference on Multimodal interfaces*. ACM, 135–142.

[8] Justine Cassell, Hannes Högni Vilhjálmsson, and Timothy Bickmore. 2001. Beat: the behavior expression animation toolkit. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 477–486.

[9] Denise Davila and Lisa Patrick. 2010. Asking the experts: What children have to say about their reading preferences. *Language Arts* 87, 3 (2010), 199.

[10] Guillaume Dubuisson Duplessis, Alexandre Pauchet, Nathalie Chaignaud, and Jean-Philippe Kotowicz. 2017. A conventional dialogue model based on dialogue patterns. *International Journal on Artificial Intelligence Tools* 26, 01 (2017), 1760009.

[11] Gunes Erkan, Arzucan Ozgur, and Dragomir R Radev. 2007. Semi-supervised classification for extracting protein interaction sentences using dependency parsing. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.

[12] Manaal Faruqui, Jesse Dodge, Sujay K Jauhar, Chris Dyer, Eduard Hovy, and Noah A Smith. 2014. Retrofitting word vectors to semantic lexicons. *arXiv preprint arXiv:1411.4166* (2014).

[13] Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Bradford Books.

[14] Katrin Fundel, Robert Küffner, and Ralf Zimmer. 2006. RelExâĂŤRelation extraction using dependency parse trees. *Bioinformatics* 23, 3 (2006), 365–371.

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

[16] Felix Hill, Antoine Bordes, Sumit Chopra, and Jason Weston. 2015. The goldilocks principle: Reading children's books with explicit memory representations. *arXiv preprint arXiv:1511.02301* (2015).

[17] Christine Howe. 2009. *Peer groups and children's development*. Vol. 14. John Wiley & Sons.

[18] Robyn Hromek and Sue Roffey. 2009. Promoting Social and Emotional Learning With Games: âĂIJItâĂŹs Fun and We Learn ThingsâĂİ. *Simulation & Gaming* 40, 5 (2009), 626–644.

[19] Peter Jansen, Niranjan Balasubramanian, Mihai Surdeanu, and Peter Clark. 2016. What's in an Explanation? Characterizing Knowledge and Inference Requirements for Elementary Science Exams. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 2956–2965.

[20] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol. 1. 687–696.

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

[22] Gary Jones, Frank E Ritter, and David J Wood. 2000. Using a cognitive architecture to examine what develops. *Psychological Science* 11, 2 (2000), 93–100.

[23] Celeste Kidd and Benjamin Y Hayden. 2015. The psychology and neuroscience of curiosity. *Neuron* 88, 3 (2015), 449–460.

[24] Dorling Kindersley. 2013. *Nursing Research Society of India*.

[25] John E Laird. 2008. Extending the Soar cognitive architecture. *Frontiers in Artificial Intelligence and Applications* 171 (2008), 224.

[26] Thomas K Landauer. 2006. *Latent semantic analysis*. Wiley Online Library.

[27] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion.. In *AAAI*, Vol. 15. 2181–2187.

[28] Markus Lindberg, Kristian Månsson, Birger Johansson, Agneta Gulz, and Christian Balkenius. 2017. Does a robot tutee increase childrenâĂŹs engagement in a learning-by-teaching situation?. In *International Conference on Intelligent Virtual Agents*. Springer, 243–246.

[29] Hugo Liu and Push Singh. 2004. ConceptNetâĂŤa practical commonsense reasoning tool-kit. *BT technology journal* 22, 4 (2004), 211–226.

[30] George Loewenstein. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin* 116, 1 (1994), 75.

[31] Marsha C Lovett, Larry Z Daily, and Lynne M Reder. 2000. A source activation theory of working memory: Cross-task prediction of performance in ACT-R. *Cognitive Systems Research* 1, 2 (2000), 99–118.

[32] Stefan Manojlovic, Katerina Gavrilo, Jan de Wit, Vassilis-Javed Khan, and Panos Markopoulos. 2016. Exploring the potential of children in crowdsourcing. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1250–1256.

[33] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[34] Wookhee Min, Joseph B Wiggins, Lydia Pezzullo, Alexandria Katarina Vail, Kristy Elizabeth Boyer, Bradford W Mott, Megan Frankosky, Eric N Wiebe, and James C Lester. 2016. Predicting Dialogue Acts for Intelligent Virtual Agents with Multimodal Student Interaction Data.. In *EDM*. 454–459.

[35] Tom Mitchell, William Cohen, Estevam Hruschka, Partha Talukdar, B Yang, J Betteridge, A Carlson, B Dalvi, M Gardner, B Kisiel, et al. 2018. Never-ending learning. *Commun. ACM* 61, 5 (2018), 103–115.

[36] Bhargavi Paranjape, Zhen Bai, and Justine Cassell. 2018. Predicting the Temporal and Social Dynamics of Curiosity in Small Group Learning. In *International Conference on Artificial Intelligence in Education*. Springer.

[37] Hae Won Park, Rinat Rosenberg-Kima, Maor Rosenberg, Goren Gordon, and Cynthia Breazeal. 2017. Growing growth mindset with a social robot peer. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 137–145.

[38] Judy M Parr and Michael AR Townsend. 2002. Environments, processes, and mechanisms in peer learning. *International journal of educational research* 37, 5 (2002), 403–423.

[39] Daven Peng and Lodovica Biagi. 2016. Children and Media in the U.S. www.statista.com/study/21298/children-and-media-in-the-us-statista-dossier/

[40] Jenefer Philp, Rebecca Adams, and Noriko Iwashita. 2013. *Peer interaction and second language learning*. Routledge.

[41] Jean Piaget. 2013. *The moral judgment of the child*. Routledge.

[42] Sameerchand Pudaruth, Sandiana Amourdon, and Joey Anseline. 2014. Automated generation of song lyrics using CFGs. In *Contemporary Computing (IC3), 2014 Seventh International Conference on*. IEEE, 613–616.

[43] Victoria J Rideout, Ulla G Foehr, and Donald F Roberts. 2010. Generation M [superscript 2]: Media in the Lives of 8-to 18-Year-Olds. *Henry J. Kaiser Family Foundation* (2010).

[44] Kimiko Ryokai, Cati Vaucelle, and Justine Cassell. 2003. Virtual peers as partners in storytelling and literacy learning. *Journal of computer assisted learning* 19, 2 (2003), 195–208.

[45] Lenhart K Schubert and Chung Hee Hwang. 2000. Episodic Logic meets Little Red Riding Hood: A comprehensive, natural representation for language understanding. *Natural language processing and knowledge representation: Language for Knowledge and Knowledge for Language* (2000), 111–174.

[46] Dorothy G Singer and Jerome L Singer. 2012. *Handbook of Children and the Media*. Sage.

[47] Tanmay Sinha, Zhen Bai, and Justine Cassell. 2017. Curious minds wonder alike: studying multimodal behavioral dynamics to design social scaffolding of curiosity. In *European Conference on Technology Enhanced Learning*. Springer, 270–285.

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

[49] Brian W Sturm. 2003. The information and reading preferences of North Carolina children. *School Library Media Research* 6 (2003).

[50] Ron Sun. 2006. The CLARION cognitive architecture: Extending cognitive modeling to social simulation. *Cognition and multi-agent interaction* (2006), 79–99.

[51] Andrea Tartaro and Justine Cassell. 2006. Authorable virtual peers for autism spectrum disorders. In *Proceedings of the Combined workshop on Language-Enabled Educational Technology and Development and Evaluation for Robust Spoken Dialogue Systems at the 17th European Conference on Artificial Intellegence*.

[52] Alexandra To, Jarrek Holmes, Elaine Fath, Eda Zhang, Geoff Kaufman, and Jessica Hammer. 2017. Modeling and Designing for Key Elements of Curiosity: Risking Failure, Valuing Questions. (2017).

[53] Marilyn A Walker, Jean E Fox Tree, Pranav Anand, Rob Abbott, and Joseph King. 2012. A Corpus for Research on Deliberation and Debate.. In *LREC*. 812–817.

[54] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes.. In *AAAI*, Vol. 14. 1112–1119.