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

Bhargavi Paranjape[1], Yubin Ge[2], Zhen Bai[1], Jessica Hammer[1], Justine Cassell[1*]
Carnegie Mellon University[1] University of Pittsburgh[2]
bvp@cs.cmu.edu, yug37@pitt.edu, {zhenb, hammerj, justine}@cs.cmu.edu

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 authority in knowledge. For a virtual agent to stimulate curiosity, it should be able to generate curiosity-eliciting verbal behaviors such as hypothesis verbalization and argumentation, in the manner that simulates peer-like cognitive and behavioral abilities. In this paper, we design and implement a virtual peer that can carry out key curiosity-eliciting science talk during a dialog-based multi-party board game. We propose a child-centered and data-driven approach to simulate the latent reasoning process of young children and age-appropriate language during open-ended game play. In particular, we use a combination of child knowledge-graph construction and child-child interaction driven modeling to generate game appropriate behaviors that are compatible with 9-14 year old children. Encouraging human evaluation of the generated behaviors and generalizability 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:
Bhargavi Paranjape[1], Yubin Ge[2], Zhen Bai[1], Jessica Hammer[1], Justine Cassell[1]. 2018. Towards Automatic Generation of Peer-Targeted Science Talk in Curiosity-Evoking Virtual Agent. In *IVA '18: International Conference on Intelligent Virtual Agents (IVA '18), November 5–8, 2018, Sydney, NSW, Australia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3267851.3267894

1 INTRODUCTION

Were you one of those children who took apart clocks to see how they work? Did you have classmates that built computers from scratch? Such information seeking behaviors are often motivated by curiosity. Following prior work, we define *curiosity* as the 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 [25]. It is one of the important social-emotional learning skill that leads to exploratory knowledge seeking [1]. Recent studies show that peer interactions exert a stronger influence on curiosity than one's own behaviors, during small group learning among 5th and 6th graders [45]. In particular, behaviors that elicit uncertainty and conceptual conflict amongst peers, such as question asking, hypothesis verbalization, argumentation and justification, tend to evoke curiosity not only in one individual, but across several group members [37].

Game play provides special opportunities to assess and support curiosity through a safe and playful environment to explore and experience uncertainty, in both individual [23] and group [50] setting. The scaffolding for "social connectedness and meaningful participation" that a game provides [20] may also makes it an engaging activity for virtual peers that exhibit the appearance and ability of a real child. Peer scaffolding has been shown to support positive development in children for curiosity [17], growth of mind [38] and social interaction [6].

In this work, we design and develop a virtual peer to scaffold curiosity in the context of a custom-designed collaborative board game called Outbreak [50]. The game incorporates the open-ended and investigative nature of curiosity by asking players to explore unknown threats through open-ended question asking and discussion, and then to decide a set of skill cards to use to conquer these threats. To engage in the game play as a competent collaborator, the virtual child needs to make real-time reasoning that processes both spontaneous peer interactions (e.g. question asking, argumentation, hypothesis verbalization) and dynamic game state updates (e.g. drawing cards). 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.

In particular, there are 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 uncertainty and knowledge gap [32], such as creating new hypotheses or arguing for alternate viewpoints. Open-ended possibilities create opportunities for others to respond to uncertainty and alternatives and may in turn lead to new knowledge gap and dissonance. Second, the virtual peer has to demonstrate equal abilities, 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 [40].

*Both first and second authors contributed equally to the paper

IVA '18, November 5–8, 2018, Sydney, NSW, Australia

ACM ISBN 978-1-4503-6013-5/18/11…$15.00
https://doi.org/10.1145/3267851.3267894

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 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. 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 encountered in the limited number of recorded game play interactions. 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, along with the constructed knowledge representation, are used to plan the content for context appropriate behaviors. We also extract language motifs from game-play corpus 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. We discuss the generalizability of the child-centric modeling approach and the behavior generation framework to other collaborative tasks that require open-ended and creative peer interactions.

The main contributions of this paper are two-fold:

(1) We present a child-centered data-driven approach to simulate peer-like cognitive and behavioral abilities during open-ended conversation.
(2) We applied the approach in developing a virtual peer that carries out peer-targeted curiosity-eliciting science talk in a multi-party game play.

2 RELATED WORK

Virtual Peers with co-equal abilities Peers are individuals of similar age, ability, knowledge, experience and social status [19, 39]. 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 [6], teachable tutees [30], and supportive companions [17, 38]. 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 [49] and Alex is a culturally-authentic virtual peer who engages a child in science talk while switching dialects [8]. 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.

Adapting Cognitive Architectures for Children Cognitive architectures are 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 [24, 33]. 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 for different developmental stages, media exposure is considered one of the main sources for children's knowledge acquisition [44], 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 [36], and previously mentioned child-child interaction in science discovery and storytelling.

Semantic Memory, Knowledge Representation and Reasoning Semantic memory stores the concepts required for reasoning by the virtual agent. Semantic memory has traditionally been represented as a semantic graph consisting of nodes (concepts or terms) and edges(relations) [29]. There are several manually curated or automatically generated knowledge bases such as WordNet [14] and ConceptNet [31] etc., that are directly incorporated into a virtual agent dialogue manager for reasoning and next utterance formulation [43]. However, these knowledge bases contain factual information and common sense associations (such as *USA-President-Trump*) that are very generic for game play and are closer to an adult's semantic memory. Distributional models of semantic memory [5] use text corpus-based co-occurrence models like LSA [28] or the more recent Word2Vec [35] to create a semantic vector space in which different concepts(terms) from the corpus reside. Such techniques have previously been used by [22, 52] to project terms into vector spaces using labeled data and supervised machine learning.

Reasoning over the knowledge graph involves making inferences based on semantic relations among various concepts. Most knowledge bases like WordNet are ontologies or taxonomies that support semantic relations such as *is a, type of, synonyms, antonyms, ways to, causes for etc.* Some works have looked at automatic relation extraction. Automated biological hypothesis generation *OpenCog* extracts relations automatically from free text [16]. Most automatic relation extraction methods like these use dependency parsing to extract underlying relations. For instance, [12] design a semi-supervised method for extracting protein-protein relations based on dependency parse trees.

Natural Language Generation The generation of natural language utterances requires the construction of a child-centered language model to capture common lexical patterns children use while talking to peers. Due to a lack of sufficient child-child data, statistical pattern-mining based approaches are a more feasible strategy. [7] extract response structures from previously annotated sequences to generate factoid questions. [11] investigate and extract dialogue patterns from human-human interactions to be used by a software agent to interact with real humans . [41] construct sentence patterns in the form of template sequences of parts-of-speech and a simple lexicon of words to populate these templates.

3 METHOD

In this work, we build a virtual peer that can generate key curiosity-inducing behaviors to elicit uncertainty and conceptual conflict in

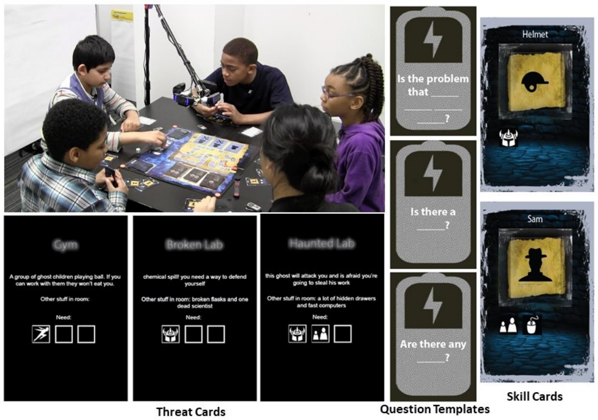

Figure 1: Outbreak game play recording and game components (Threat cards, question templates, Skill cards)

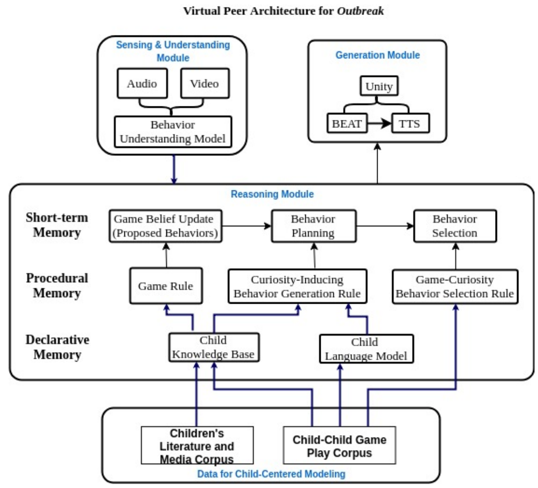

Figure 2: Data-driven System Architecture of the Outbreak Virtual Peer

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 [50]. 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, to uncover unknown threats. Players then 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 (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. An example game scenario is depicted in Figure 1

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 local 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 cover all rooms and reach the end of the game. 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 (3630 clauses in total).

We used two human annotators to annotate every clause in our corpus for three key verbal behaviors that have been correlated with increased curiosity - Justification, Argumentation and Hypothesis Verbalization [37]. *Justification* refers to showing that somebody or something is right or reasonable. *Argument* is an exchange of diverging or opposite views among multiple people, including providing reasons to change people's positions. *Hypothesis Verbalization* refers to expressing different possibilities or theories that explain something that happened, often by giving a relation between two or more variables. Inter rater reliability (Krippendorf's alpha) for each behavior annotation 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 clauses.

We describe the cognitive architecture of the virtual peer (depicted in Figure 2), that refers to key components of general cognitive architecture proposed in ACT-R [4], Soar [27] and CLARION [48] 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 verbal behaviors to engage in game play and provide open-ended possibilities for evoking curiosity ; (3) behavior generation - realize the verbal behaviors through text-to-speech (TTS) using Amazon Polly, and associated non-verbal behaviors using Behavior Expression Animation Toolkit (BEAT) [9] 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 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. Semantic memory encodes word meanings, facts, concepts, and general world knowledge that agent uses to reason about the game, while procedural memory maintains rules required for planning the content, natural language form and selection criteria of different verbal behaviors.

Processing Unit. The processing unit refers to the key reasoning steps over the components of the cognitive architecture. 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 skills that associate with the keyword "robot" (e.g. fight, be friend, block)
(2) **Behavior Content Planning** - Plans for the content of game-play behaviors and curiosity-inducing verbal behaviors, according to the game belief updates. For example, in the discussion phase, since the game master has clarified that the room has a robot, 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, children may be more likely to make arguments following another person's suggestion than new hypotheses, so the generated argument behavior is chosen more often than the hypothesis.

The game belief update and behavior planning are driven by a child knowledge graph, a distributional representation of semantic memory that is generated using 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 graph in our framework) connects entities using relations that represent a shared encoding of their meaning. A knowledge graph encodes the relative meaning of concepts accrued by humans from repeated episodic experiences. Hence, we begin with the construction of a general child-centered knowledge graph and develop strategies to traverse this graph.

3.1.1 Collecting Auxiliary Dataset. Our existing child-child gameplay corpus 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 that contains the knowledge that real children of the target age group are exposed to [44] while reading books and watching videos. This data acts as proxy for knowledge children acquire from day-to-day experiences and is able to encode conceptual associations that a human peer is likely to make in game play. We reviewed articles and surveys about children's media usage patterns and preferences for specific genres. [42] 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 and entertainment. Besides, e-books also attract a substantial number of children. Along the dimension of genre, we explored the kinds of information that children are most interested in. [46] 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. [10] review the authorized reading materials. Their study reveals that boys and girls equally like fictions which contain horror, humor and adventure. 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)[1]. We also included the children's literature corpus released by Facebook consisting of fairy tales and story books from Project Gutenberg [18]. The size of the text corpus collected was 100 MB.

3.1.2 Vector Spaces as Child Knowledge Representation. We train a distributed word embedding model (Word2Vec [35]) 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.

3.1.3 Task-Specific Transformation of Knowledge.

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*. 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 associations between skills and keywords extracted from the Outbreak game play corpus as a seed. We extract new potential keywords for a skill s as follows:

$$argmin[\alpha(\| x - x_c \|^2) + \beta \| x - x_s \|^2] \quad (1)$$

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

[1]Download links to various sources we collected data from : https://tinyurl.com/childmediacorpus

Table 1: The Top 5 Keywords For a Part of Skill Words

Skill	Fight	Love Animals	Hack	See
Keywords	Enemy Attack Terrible Deadly Monster ...	Animal Dog Creature Cat Bird ...	Robot Machine Computer Software Automatic ...	Escape Light Dark Get away Bright ...

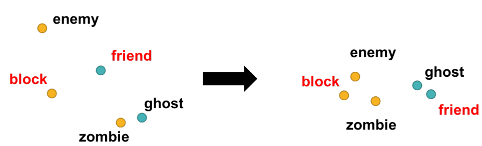

Figure 3: Retrofitting Knowledge Graph to be more game-specific: The dots with the same color are mapped to the same skill, the words in red are skill words and the rest are keywords

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 1 shows new top 5 keywords from ~400 keywords. This process results in links between potentially 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 a specific skill card down when a new keyword is encountered in future games. The virtual agent would suggest one of the skills most strongly mapped to that keyword by Equation 1, so the 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 card *see*, a question like "Does the room look dark?" might be asked. These keyword-skill associations enable the virtual agent to engage in basic game play with real children. We evaluate these associations manually, which we describe in Section 4.

Adapting Knowledge Graph 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 graph 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 [3]. 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 [13] 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, like the skills in this case.

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 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 parsing based syntactic analysis of the game-play corpus to discover common syntactic structures children use when they display these three verbal behaviors. Dependency parsing breaks sentences down to their syntactic tree-structures. However, the size of our child-child data is limited and curiosity-stimulating behaviors are sparse, resulting in infrequent patterns. Once again, we turned to auxiliary datasets that have been created for the verbal behaviors we wanted to generate. For example, AI2 Elementary and Middle School Science Questions corpus was manually annotated for *Justification* [21] and the Internet Argument Corpus has been annotated for 3079 *Argument* instances [51]. Dependency Parsing was applied to both datasets. We find that the dependency relations of *amod (adjectival modifier of a noun phrase)* and *dobj(direct object of a verb phrase)* rank in the top 7 among 40 dependency relations. The other top relation types such as *det(determiners)* and *punc(punctuations)* are essential grammatical constructs present in most sentences but play no 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 [37]. [26] claim that a hypothesis is a conjectural statement that encodes the relation between two or more variables, so *amod* and *dobj* 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 two relations from the collected children's corpus(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 made by a real child.

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 retrofitted child knowledge graph. The most generalizable search strategy we develop is the bottom-up strategy, shown in Figure 4: Given a keyword and skill, we find

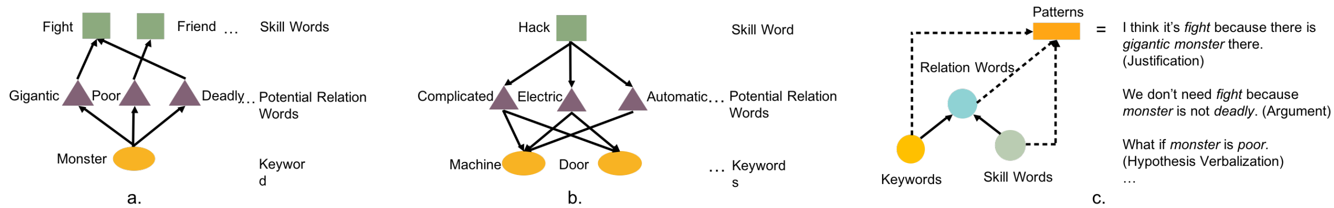

Figure 4: a. The Figure of Bottom Up Strategy in an Example; b. The Figure of Top Down Strategy in an Example; c. The Figure Showing the Overall Strategy for NLG

terms related by the *amod, dobj* relations from our general children's corpus to the keyword that are semantically similar to the skill. We start from the keyword, and iterating over most frequent relation words, we calculate their semantical similarities with the skill using the following formula.

$$\alpha * sim(relation, skill) + \beta * sim(keyword, skill) \quad (2)$$

,where sim(x, y) calculates the similarity of vectors for words x and y in the retrofitted word vector space, and α and β are used for controlling the relative strength of each term in 2. Figure 4 shows the result of combining the strategy with knowledge graph for final utterance generation.

For **Justification** we pick a keyword in the known game context and use the knowledge graph to find an appropriate skill as before. A relation word is generated to support this association 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 Equation 2 and a top-scoring term is chosen to complete the justification. A possible generated sentence can be: *We need fight because there is gigantic monster there.*

Argument often happens when one gives reasons to deviate from someone else's ideas. We observe from child-child game play data that children often argue against a skill suggested by other players. For example, if *fight* is suggested by a player for the keyword *monster*, we score and rank related words as before. A possible generated argument can be: *No fight because the monster isn't deadly.* Another approach for argument generation is to argue for another skill for a mentioned keyword. We start from a mentioned keyword and choose the next closest skill from the knowledge graph. For example, the next relevant skill word for *monster* is *friend*, and the generated argument can be: *No, the monster can be hurt, so we need friend.*

Hypothesis verbalization helps in expressing new possible relations between two entities using conjectural statements. Using the bottom-up strategy, we propose an alternative attribute about a keyword as a hypothesis. For instance, the keyword *monster* is usually related to *fight*, but to encourage other players to explore alternate skills, we can use a word related to the skill *friend* and make the following hypothesis: *But what if the monster is poor!!.*

An important observation about children's game play is that they don't always use information they acquire during the Q&A phase. Instead, they may consider the skill cards they possess in that game round. For this, we develop a top-down search strategy (refer Figure 4). For each skill, we extract a pair of *amod* and *dobj* related words

Table 2: Examples of Generated Child Language 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*.

that are closely associated with the skill. This strategy is especially helpful for hypothesis verbalization, which can be generated based on initiation from the skills without game context information. For instance, for the virtual child to use the *Hack* skill, the top-down strategy can output a sentence like: *There might be an automatic machine in the room.* Such a sentence promotes other children to think more about the hypothesis or generate other plausible hypotheses.

3.2.3 Natural language utterance generation. Verbal behaviors are planned with the selection of appropriate skill, keyword and relation word. To generate the final agent utterance, we create sentence patterns where the chosen words are filled in. To obtain templates that closely resemble children's speech during gameplay, another dependency parsing was performed on Outbreak game play data, and for each verbal behavior we pick the top 10 common syntactic structures and extract the corresponding natural language patterns 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. Here too, we let the child-child interaction data inform the language of the virtual agent to make it more believable and peer-like. Table 2 shows the top templates extracted for the three behaviors.

4 EVALUATION

Game Belief Update Evaluation. Precision and Recall are adopted to evaluate the newly extracted keyword-skill associations. For precision, we generated top 50 potential keywords for 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. We disambiguate using majority vote among human annotators and leave out cases where even human annotators disagree. We calculated precision as the proportion of associations that the computational model retrieves that are deemed reasonable and get a score of 0.83. Krippendorff's alpha for annotators' inter-rater reliability is 0.86. For Recall, we picked 200 words from the larger peer-targeted media corpus and asked annotators to map each word to one or more of the 7 skills. We again use majority vote as the ground truth for reasonable associations. Krippendorff's alpha was 0.86. Our computational model is used to find the corresponding highest ranked skills only for associations marked reasonable, to measure if one or more of them are correctly retrieved. The recall was 0.68.

Curiosity-related Verbal Behaviors Generation 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 clauses for three verbal behaviors and also mixed them with 100 randomly picked clauses from child-child game play corpus. We asked the annotators to classify the utterances into the 3 categories based on the generated utterance and the available context. The computed accuracy of classification over all three behaviors is 0.77 and the Krippendorff's alpha for agreement among annotators is over 0.7 for all three behaviors.

5 DISCUSSION

Knowledge Graph Construction. We validate that the generated skill-keyword associations 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 text 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 age-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 all game-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 [2]. In recent work, agents are using distributional semantic memory for small reasoning tasks - robots that understand analogies in human instructions [47] or agents that detect behavioral affordances such as objects that can be grasped, drunk, worn, etc [15].

Curiosity-Inducing Behavior Generation. We automate the agent's curiosity-related verbal behavior generation for a limited number of syntactic relations(*dobj, amod*) between entities. The promising evaluation of the generated curiosity-related behaviors for game-context appropriateness supports 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. Our method supports syntactic and semantic relations to generate context relevant and coherent sentences. An important future work is to understand causal relations between concepts and the pragmatics of human conversation that are 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. Procedural memory 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 children. This will be done in order to validate if curiosity-inducing behaviors selected and generated by the virtual peer can induce 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 semantic 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' evaluations of children's reasoning and language. A thorough evaluation of the relative success of this child-centered data-driven technique should involve children of the target age. 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 in the HCI community. For example, Manojlovic et al. [34] found that joint tasks solved by 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. This can further fine-grained incremental development of virtual peer technology rather than final user studies with children.

6 CONCLUSION

Peer interaction may serve as special stimulus in inducing curiosity by facilitating increased uncertainty and conceptual conflicts and collaborative games provide an ideal setting to express uncertainty. In this work, we build a virtual peer agent that can elicit curiosity stimulating behaviors while engaging in a discussion based board game. We develop and implement a behavior generation framework to realize peer-targeted curiosity-inducing behaviors such as hypothesis verbalization, argumentation and justification during game play. We create a child-centered knowledge graph and make game-specific adaptations to the induced graph. We use child-child interaction driven behavior and language modeling to generate spontaneous and context-appropriate verbal behaviors. Promising intrinsic evaluations of the generated behaviors for game-context appropriateness and generalizability of the generation framework opens up encouraging new directions in virtual peer modeling for open-ended game play and science reasoning tasks. In future work, we intend to carry out user testing with children to validate if generated behaviors can stimulate curiosity in peers.

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] 2018. Enhancing Virtual Agents with Structured Knowledge. https://www.econtext.ai/wp-content/econtextfiles/eContext-VA-Whitepaper.pdf
[3] John R Anderson. 1983. A spreading activation theory of memory. *Journal of verbal learning and verbal behavior* 22, 3 (1983), 261–295.
[4] 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.
[5] Marco Baroni and Alessandro Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics* 36, 4 (2010), 673–721.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] Denise Davila and Lisa Patrick. 2010. Asking the experts: What children have to say about their reading preferences. *Language Arts* 87, 3 (2010), 199.
[11] 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.
[12] 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)*.
[13] 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).
[14] Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Bradford Books.
[15] Nancy Fulda, Daniel Ricks, Ben Murdoch, and David Wingate. 2017. What can you do with a rock? Affordance extraction via word embeddings. *arXiv preprint arXiv:1703.03429* (2017).
[16] Katrin Fundel, Robert Küffner, and Ralf Zimmer. 2006. RelExâĂŤRelation extraction using dependency parse trees. *Bioinformatics* 23, 3 (2006), 365–371.
[17] 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.
[18] 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).
[19] Christine Howe. 2009. *Peer groups and children's development*. Vol. 14. John Wiley & Sons.
[20] Robyn Hromek and Sue Roffey. 2009. Promoting Social and Emotional Learning With Games: âĂIJIt's Fun and We Learn ThingsâĂİ. *Simulation & Gaming* 40, 5 (2009), 626–644.
[21] 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.
[22] 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.
[23] Jamie Jirout and David Klahr. 2012. Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review* 32, 2 (2012), 125–160.
[24] 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.
[25] Celeste Kidd and Benjamin Y Hayden. 2015. The psychology and neuroscience of curiosity. *Neuron* 88, 3 (2015), 449–460.
[26] Dorling Kindersley. 2013. *Nursing Research Society of India*.
[27] John E Laird. 2008. Extending the Soar cognitive architecture. *Frontiers in Artificial Intelligence and Applications* 171 (2008), 224.
[28] Thomas K Landauer. 2006. *Latent semantic analysis*. Wiley Online Library.
[29] 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.
[30] Markus Lindberg, Kristian Månsson, Birger Johansson, Agneta Gulz, and Christian Balkenius. 2017. Does a robot tutee increase children's engagement in a learning-by-teaching situation?. In *International Conference on Intelligent Virtual Agents*. Springer, 243–246.
[31] Hugo Liu and Push Singh. 2004. ConceptNet a practical commonsense reasoning tool-kit. *BT technology journal* 22, 4 (2004), 211–226.
[32] George Loewenstein. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin* 116, 1 (1994), 75.
[33] 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.
[34] 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.
[35] 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.
[36] 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.
[37] 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, 420–435.
[38] 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.
[39] Jenefer Philp, Rebecca Adams, and Noriko Iwashita. 2013. *Peer interaction and second language learning*. Routledge.
[40] Jean Piaget. 2013. *The moral judgment of the child*. Routledge.
[41] 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.
[42] 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).
[43] 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.
[44] Dorothy G Singer and Jerome L Singer. 2012. *Handbook of Children and the Media*. Sage.
[45] 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.
[46] Brian W Sturm. 2003. The information and reading preferences of North Carolina children. *School Library Media Research* 6 (2003).
[47] Douglas Summers-Stay and Dandan Li. 2017. Analogical Reasoning with Knowledge-based Embeddings. (2017).
[48] Ron Sun. 2006. The CLARION cognitive architecture: Extending cognitive modeling to social simulation. *Cognition and multi-agent interaction* (2006), 79–99.
[49] 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*.
[50] 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).
[51] 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.
[52] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes.. In *AAAI*, Vol. 14. 1112–1119.