A Novel Multimodal Approach for Studying the Dynamics of Curiosity in Group Learning

Tanmay Sinha, Student Member, SIGCHI, Zhen Bai, Member, SIGCHI, and Justine Cassell, Fellow, ACM

Abstract—Curiosity is a vital metacognitive skill in educational contexts. Yet, little is known about how social factors influence curiosity in group work. We argue that curiosity is evoked not only through individual, but also interpersonal activities, and present what we believe to be the first theoretical framework that articulates an integrated socio-cognitive account of curiosity based on literature spanning psychology, learning sciences and group dynamics, along with empirical observation of small-group science activity across three formal and informal learning contexts. We make a bipartite distinction between individual and interpersonal functions that contribute to curiosity, and multimodal behaviors that fulfill these functions. We validate the proposed framework by leveraging a longitudinal latent variable modeling approach. Findings confirm positive predictive relationship of the latent variables of individual and interpersonal functions on curiosity, with the interpersonal functions exercising a comparatively stronger influence. Prominent behavioral realizations of these functions are also discovered in a data-driven way. We further extend the proposed theoretical framework by outlining a comprehensive set of strategies and tactics that can be incorporated in learning technologies to support putative functions of curiosity. Overall, this work is a step towards designing learning technologies that can recognize and evoke moment-by-moment curiosity during learning in social contexts. The underlying rationale is applicable more generally for modeling and developing computer support for other metacognitive and socio-emotional skills as well.

Index Terms—Curiosity, Learning in Social Contexts, Multimodal Human Behavior Analyses, Scaffolding

1 INTRODUCTION

Curiosity pertains to the strong desire to learn or know more about something or someone, and is an important metacognitive skill to prepare students for lifelong learning ([1], [2]). Traditional accounts of curiosity in psychology and neuroscience focus on how it can be evoked via underlying mechanisms such as novelty (features of a stimulus that have not yet been encountered), surprise (violation of expectations), conceptual conflict (existence of multiple incompatible pieces of information), uncertainty (the state of being uncertain), and anticipation of new knowledge ([3], [4]). These knowledge seeking experiences create positive impact on students beliefs about their competence in mastering scientific processes, in turn promoting greater breadth and depth of information exploration [5]. These theories have inspired the development of several computer systems aiming to facilitate task performance via enhancing an individual’s curiosity (e.g. [5], [6], [7]), simulating human-like curiosity in autonomous agents [8], and aiding in game theory development [9]. Evoking curiosity in these systems mainly focuses on directing an individual to a specific new knowledge component, followed by facilitating knowledge acquisition through exploration. Such a linear approach largely ignores the how learning is influenced when working in social contexts. Here, a child’s intrinsic motivation, exploratory behaviors, and subsequent learning outcomes may be informed not only by materials available to the child, but also the active work of other children, social and cultural environment, and presence of facilitators ([10], [11]). For example, an expression of uncertainty or of a hypothesis about a phenomenon made by one child may cause peers to realize that they too are uncertain about that phenomenon, and therefore initiate working together to overcome the cause of uncertainty, in turn positively impacting their curiosity [12]. While prior literature has extensively studied the intrapersonal origins of curiosity, there seems to be very little prior work on how social factors contribute to moment by moment changes in an individual’s curiosity when learning in social contexts (except for rare exceptions such as [13] that primarily focused on coarse-grained study of adult-child interaction).

As learning in small group becomes prevalent in today’s classrooms [10], it is critical to understand curiosity beyond the individual level to an integrated knowledge-seeking phenomenon shaped by social environment. Embodied Conversational Agents (ECAs) have demonstrated special capacity in supporting learning and collaborative skills for young children [14]. Knowing how social factors influence curiosity allows researchers to design ECAs and other learning technologies to support curiosity-driven learning before children naturally support each other. To address the above goal, we first propose an integrated socio-cognitive account of curiosity based on literature spanning psychology, learning sciences and group dynamics, and empirical observation of an informal learning environment.

We make a bipartite distinction between putative functions that contribute to curiosity, and multimodal behav-

---

- T. Sinha is with Institute of Learning Sciences and Higher Education, ETH Zurich, Switzerland. E-mail: see http://www.tanmay.org/tanmaywebsite
- Z. Bai is with Human Computer Interaction Institute, School of Computer Science, Carnegie Mellon University, USA. E-mail: zhenb@cs.cmu.edu
- J. Cassell is with Language Technologies Institute, School of Computer Science, Carnegie Mellon University, USA. E-mail: justine@cs.cmu.edu

Manuscript received XX; revised XX.
tors that fulfill these functions. These functions comprise (i) “knowledge identification and acquisition (helps humans realize that there is something they desire to know, and leads to acquisition of the desired new knowledge), and (ii) “knowledge intensification” (escalates the process of knowledge identification or acquisition by providing favorable environment, attitude etc) - at individual and interpersonal level. Second, we perform a statistical validation of this theoretical framework to illuminate predictive relationships between multimodal behaviors, functions (latent variables because they cannot be directly observed) and ground truth curiosity (as judged by naive annotators). A longitudinal latent variable modeling approach is used to explicitly account for group structure and differentiate fine-grained behavioral variations across time.

The main contributions of this work are three-fold: First, it begins to fill the research gap of how social factors, especially interpersonal peer dynamics in group work, influence curiosity (section 6). Second, the model is designed to lay a theoretical foundation to inform the design of learning technologies, a virtual peer in the current study, that employ pedagogical strategies to evoke and maintain curiosity in social environments (section 7). Findings derived from the current analyses of human-human interaction can be informative in guiding the design of human-agent interaction. Third, at the methodological level, our research (i) introduces novel approaches for collecting rich multimodal data in group settings (section 4.1), which is key to making fine-grained behavioral inferences, (ii) advances use of crowdsourcing platforms for efficient ground truth annotation, which is important in human behavior analysis for educational research and beyond (section 5.1), (iii) provides a rigorous and reproducible semi-automatic behavior annotation approach, which combines complementary strengths of state-of-the-art machine learning methods and advantages of human judgment (section 5.2, 5.3 and 5.4).

In what follows next, Section 2 discusses relevant related work. Section 3 begins describing our combination of theory-driven and data-driven process for development of the theoretical framework of curiosity, in particular, outlining the putative underlying mechanisms of curiosity in group work and associated multimodal behaviors. Section 4 discusses data collection across multiple study contexts. Section 5 describes the annotation of ground truth curiosity, verbal and nonverbal behaviors, along with turn taking metrics. Section 6 discusses empirical validation of the theoretical framework of curiosity, with full technical description along with results and discussion of the model fit to our corpus. In section 7, we describe implications of this work for learning technology design by outlining a comprehensive set of strategies and tactics to foster curiosity in learning in social contexts. Finally, we end with limitations, future research directions, and conclusion in sections 8 and 9.

2 RELATED WORK

For clarity, we divide related work into four subsections. We first discuss prior theoretical stances on curiosity from the psychology literature, and subsequently discuss the importance of social factors in understanding curiosity from the group dynamics literature. We then outline how curiosity as a construct has been perceived in the learning sciences, and end with a brief overview of existing computational modeling approaches for curiosity.

2.1 Curiosity in the Psychology Literature

Researchers in psychology describe curiosity as a psychological and behavioral state that “responds to an inconsistency or a gap in knowledge” [15], and raises “feelings of mystery, of strangeness, and of wonder” [16]. Like hunger and thirst, curiosity is considered a critical internal drive for human beings to explore our environment, acquire knowledge and learn skills. It is generally described as an intrinsically motivated desire, passion or appetite for information, knowledge and learning [17]. The incongruity theory argues that curiosity is the human tendency to make sense of the world on observing violated expectation ([18], [19]). Conflict arousal theory considers curiosity as a drive stimulated by psychological conflict that leads to simultaneous occurrence of incompatible response tendencies. Main determinant factors for psychological conflict include perceptual factors (novelty, surprisingness, incongruity, uncertainty, complexity) and epistemic factors (doubt, perplex, contradiction, conceptual incongruity, confusion and irrelevance) [20]. The information-gap theory proposes that curiosity is raised when people attend to a gap in their knowledge, and the intensity of curiosity depends on importance, salience and surprisingness of the desired information ([3], [21], [22]). In addition to unwanted feelings caused by deprivation of knowledge, curiosity is also conceived as desirable feelings associated with the anticipation of acquiring new knowledge. It has been posited to intrinsically motivate positive affect and receptivity to new experience, task absorption, and active exploration of topics with interests ([11], [23]).

Curiosity has generally been divided into trait and state curiosity in prior work. Trait curiosity refers to the internal capacity of individuals to experience curiosity, whereas state curiosity refers to curiosity raised under a particular circumstance [20]. Common measurement approaches for curiosity include questionnaire-based self-report methods such as Melbourne Curiosity Inventory [24], State-trait Curiosity Inventory [25], and epistemic curiosity items [26]. Compared to these self-report measures, Jirout and Klahr suggested that it is more suitable to use behavior-based evidence to measure curiosity for young children as they lack reading and comprehension skills [3]. Previous literature reveals several behavior cues for curiosity including physical exploration such as orientation, locomotion and manipulation towards objects of interest, epistemic investigation such as question asking, experimentation, description, reasoning about observed phenomenon, as well as demonstrative expressions of surprise, excitement, wonder, confusion and attentiveness ([20], [27], [28], [29]). Existing behavior-based measures of children’s curiosity mainly rely on physical exploration and simple verbal behaviors such as question asking and commenting (for e.g. [30], [31], [32], [33]). There is, however, lack of rigorous operational measures of state curiosity that incorporate comprehensive verbal and nonverbal behaviors displayed in real-time interaction, and those that “have potential to suggest practical methods of stimulating curiosity in the broader population”, but are especially challenging due to the transient characteristics of curiosity [21].
2.2 Curiosity in the Group Dynamics Literature

Social accounts for curiosity remain largely unexplored in the psychology literature, and relevant research primarily focuses on parent-child or teacher-student interactions ([13], [34]), instead of peer-peer interaction. Knowledge dissonance, social comparison, risk-taking and social information-seeking are four main social factors that are tightly related to curiosity and exploration. We describe each of these four factors below.

First, compared to teacher-led learning where the teacher holds higher power/status in terms of knowledge possession, peer-peer learning is more likely to result in challenging different opinions or ideas from each other and active resolution of such knowledge incongruity or dissonance [35], which is one of the main sources of curiosity [36]. Controversy, instantiated as conflict or disagreement, has also been identified as one of the social causes of dissonance - the simultaneous existence of cognitions that in one way or another do not fit together. When individuals working in a small group experience dissonance, they might work towards reducing or eliminating dissonance via different means. One important means is to make an attempt to seek additional social support for the held opinion by emphasizing its importance, lucrativeness, etc., in turn potentially triggering curiosity and interest from group members [35].

Second, through social comparison, students are more eager to evaluate the correctness of one’s own opinion via group discussion, when knowing their peers’ opinion, compared to knowing an expert’s opinion [37]. When individuals face a question with no clear solution and they cannot reduce the uncertainty by consulting objective sources of information, they turn to views endorsed by others in the group and evaluate the accuracy of their beliefs by comparing themselves to others [35]. In addition, cognitive and affective changes are more likely when observing others who are perceived as friends or similar to the observer [10]. Also, students are more likely to actively seek information and solutions when their own uncertainty is shared or at least considered as warranted, reasonable, or legitimate by their peers [12]. Such joint hardship [38] is likely to impact group member’s behavior positively, due to the trigger it provides for engaging in cooperative/joint effort to overcome the obstacle by reasoning or physical exploration.

Third, risk-taking, or the willingness to expose oneself to failure in pursuit of a desirable goal, influences individuals’ pursuit of desirable knowledge, as knowledge acquisition inevitably involves failure. Members in a small group tend to make riskier decisions under group influence compared to when making decisions alone, because of spreading responsibility and a decreased feeling of personal responsibility [35]. This, in turn provides a psychologically safe environment [39] to engage in free exploration, without excessive concern about others’ reaction to actions that have the potential for embarrassment or threat.

Fourth, social information seeking, or the general interest in gaining new social information (how other people behave, act and feel) promotes acceptance (a non-evaluative feeling and an unconditional positive regard towards another) [38] and creates mutually shared cognition in the group [40]. It creates space for group members to learn from others’ preferences and viewpoints, know that there are different viewpoints and accept the existence of alternative viewpoints as legitimate by perhaps considering them from their own viewpoints [41]. This increased group member familiarity and knowledge awareness can increase willingness to work jointly and entail consideration of more alternatives, and thereby the nature of communication itself and problem solutions may become more creative.

In this study, we are motivated to fill the research gap that exists in social contexts by incorporating theories from peer learning and group dynamics, as well as seeking pedagogical strategies in fostering curiosity from learning theories that intersect with curiosity and scientific inquiry. We take a socio-cognitive view [42] of curiosity, wherein we acknowledge social influences, but also try to isolate the individual mind as a cognitive unit of analysis by controlling for these external influences.

2.3 Curiosity in the Learning Sciences Literature

Compared to extensive literature on curiosity in psychology, curiosity has been scarcely discussed in the learning sciences. The roots can be traced back to [43], who differentiated extensive curiosity (that serves to widen a learner’s interest) and particular curiosity (that serves to help a learner acquire detailed knowledge). [44] tied these two notions of curiosity to the literature on knowledge awareness in collaborative learning settings. They posited that when a learner’s activities are intended towards the “same knowledge” that their peer is looking at, discussing or changing, particular curiosity is excited. On the contrary, when a learner’s activities are intended towards “different knowledge” than their peer, extensive curiosity can be satisfied and collaboration possibility is enhanced.

Curiosity has also been discussed under the umbrella of intrinsic motivation ([45], [46]). Intrinsically motivated learners derive pleasure from the task itself, while learners with extrinsic motivation rely on external rewards. [47] considers curiosity as one of the motivational aspects in the design of learning technologies, and discusses surprising students as a central instructional tactic to arouse their curiosity and lead them to explore new areas of the subject for constructing coherent explanations. More recently, curiosity has also been investigated as one of the seven dimensions of the construct of “learning power” [48], which refers to a form of consciousness, or critical subjectivity which leads to growth. Learners who are critically curious adopt deep rather than surface learning strategies [49] to get to the bottom of things, and are less likely to accept what they are told uncritically. Such learners enjoy asking questions and are more willing to reveal their questions and uncertainties.

It is important to note, however that, the focus of the learning sciences literature has been fundamentally cognitive, whereas we seek to understand the social scaffolding of curiosity along with its cognitive roots.

In artificial intelligence in education (AIED), several intelligent tutoring systems investigate adjusting teaching strategies in response to students real-time learning activities [50], [51]. There is, however, little research on how dynamics of social interaction may influence students intrinsic learning motivation. Some emerging applications of curiosity can be seen in development of pedagogical
agents. For e.g., [52] found that a curious peer (that keeps questioning) can problematize the interaction, and direct a human learner’s attention to spot contradiction in their knowledge structure, thereby inducing curiosity. [53] also discovered that if a computer agent displays curiosity by pro-actively responding to novel, complex and conflicting stimuli, it can discover interesting learning concepts. Furthermore, interaction with such a computer agent was shown to lead human learners to explore more in the learning environment, thereby retaining their attention in the task and enhancing learning outcomes. The underlying model of interpersonal influence that is common to both these research strands in AIED is “modeling”, where the assumption of human learners spontaneously picking up on social cues of the “curious” agent is made. Moreover, curiosity is very narrowly defined, while we aim to develop a more nuanced understanding of the construct, without directly equating behaviors to curiosity, or, relying solely on a single theoretical lens of looking at curiosity.

2.4 Existing Computational Models of Curiosity

In general, curiosity has been computationally modeled in prior literature [54] using an appraisal process where the incoming stimuli is first evaluated for its potential to provide an appropriate stimulation level, which is then followed by mapping the stimulation level to a non-linear emotion curve called the Wundt curve [55] to derive the curiosity level. The Wundt curve postulates that too little stimulation results in boredom, too much stimulation results in anxiety, and only optimal stimulation can result in curiosity. Across the two stages (evaluation of stimulation level and evaluation of curiosity level), prior research has also made fine-grained distinctions between one or more collative variables of curiosity such as novelty, surprise, uncertainty, complexity etc. For instance, the appraisal of novelty can be quantified by considering states that produce high empirical prediction errors (meaning those that haven’t been recently explored), while the appraisal of surprise can be quantified using Bayesian inference as a difference between a prior and posterior world model. Machine learning approaches such as fuzzy cognitive maps [56], partially observable Markov decision processes [5], self-organizing maps [57], active model babbling [58] etc have been deployed to model curiosity and information seeking.

Our sense of computational modeling of curiosity differs from the above cited literature, since we focus not just on perceptual, but also the epistemic dimension of curiosity. In addition, we intend to study learning in social contexts, and utilize embodied conversational interaction as a means to sense, reason and respond to varying curiosity levels of children working in a group. Towards this end, in this work, we propose a new theoretical framework (comprising behaviors, functions, strategies and tactics) that can be made computational by implementation of sensing, reasoning and responding modules, as we have successfully done in other prior work ([59], [60]). This is unlike goals of prior computational models of curiosity.

3 Theoretical Framework Development

We define curiosity-driven learning in social contexts to be situations where particular forms of interaction among people trigger or facilitate salient behaviors that are associated with high curiosity. To explore this notion further and initiate development of an integrated psychological and social framework of curiosity, we used a combination of theory-driven and data-driven approaches. The motivation stemmed from a dual perspective to understand curiosity - the first perspective derived from analysis of human-human interaction and reading of social science literature, while the second perspective derived from our efforts toward implementation of a learning environment and reading of the learning science literature. This led us to describe: (i) a set of putative functions that contribute to curiosity, and (ii) multimodal behaviors that provide evidence for potential presence of an individual’s curiosity in the current time-interval because of their fulfillment of these functions.

From the theory-driven perspective, we conducted several iterations of literature review with a gradual shift from individual- to interpersonal-level curiosity. Conventional perspectives on curiosity at the individual level enlightened a concise group of putative underlying functions, through which certain observable behaviors were posited to contribute to the expression and prediction of curiosity. Due to lack of research on the social influence of curiosity, we expanded our literature review to the field of learning sciences. This not only strengthened our understanding of individual curiosity in view of productive interactions in open-ended hands-on learning contexts, but also opened doors for investigation of collaborative learning behaviors that could potentially contain seeds of curiosity. Looking through the lens of learning sciences shed light on constructivism perspective of curiosity, where psychological factors of curiosity might be evoked or facilitated during collaborative construction of knowledge. Leveraging the bridge of the study of collaboration from the learning sciences perspective, we further extended the literature review to the social psychology of group dynamics, with an intent to better understand the nature of social influence in groups in our case, in learning groups. This exploration led us to seek explanations of how direct and indirect peer influence enhances or impedes curiosity in group work. Informed by the above theoretical lenses, we developed a comprehensive set of observable behaviors and putative functions that might contribute to curiosity.

From the data-driven perspective, we carried out extended empirical observation of small groups of 9-14 year old children engaged in hands-on learning activities across lab, science and STEAM class, and informal learning environments (detailed description see Section 4). We used the thematic analysis method to conduct preliminary qualitative analyses in a rich and exploratory manner based on field note, video and audio data collected during the observations [61]. We first identified a list of curiosity related individual and interpersonal behaviors during small group science learning based on our initial literature review. We then conducted qualitative analysis based on, but not constrained by, the initial behavior list, with the purpose of empirically validating and extending curiosity-related behaviors while minimizing confirmation bias. During this process, we developed thematic charts (Fig. 1) that established links between curiosity and peer-peer interaction to seek potential causal factors and signals of curiosity. These thematic charts
led to extended literature review from the learning sciences and social psychology of group dynamics, along with new iterations of qualitative analyses to obtain instances of peer interaction behaviors in our corpus that provided in-depth empirical evidence for the proposed framework.

The combination of theory-driven and data-driven approach described above allowed us to (i) obtain familiarity to describe and explain the phenomenon of curiosity in small group learning contexts; (ii) develop the integrated theoretical framework of curiosity with in-depth empirical evidence; (iii) form relevant hypotheses based on the empirical exploration for the follow-up quantitative validation of the theoretical framework. In sum, understanding of state-of-the-art science and STEAM learning environments and curricula enables promising potential for the theoretical framework of curiosity and corresponding learning technologies to be generally applicable across various formal and informal learning contexts.

3.1 Putative Functions Contributing to Curiosity

The iterative process described above led to emergence of three function groups at the individual and interpersonal level. Since curiosity has traditionally been described as an inherently individual and stable disposition toward seeking novelty and approaching unfamiliar stimuli, we first outline individual aspects of curiosity for each function. In addition, we also then flesh out interpersonal aspects of curiosity for every function. Note that each of these functions can be realized in several different behavioral forms.

We call the first function group Knowledge Identification. As curiosity arises from a strong desire to obtain new knowledge that is missing or doesn’t match with one’s current beliefs, a critical precondition of this desire is to realize the existence of such knowledge. At an individual level, knowledge identification contributes to curiosity by increasing awareness of gaps in knowledge [21], as well as highlighting relationships with related or existing knowledge in order to assimilate new information [62]. Furthermore, exposure to novel and complex stimuli can raise uncertainty, subsequently resulting in conceptual conflict [20], [36]. At an interpersonal level, knowledge identification contributes to curiosity by developing awareness of somebody else in the group having conflicting beliefs [20] and awareness of the knowledge they possess [44], so that a shared conception of the problem can be developed [40].

We call the second function group Knowledge Acquisition. This is because knowledge seeking behaviors driven by curiosity not only contribute to the satisfaction of the initial desire for knowledge, but also potentially lead to further identification of new knowledge. For example, question asking may help close one’s knowledge gap by acquiring desired information from another group member. Depending on the response received, however, it may also lead to escalated uncertainty or conceptual conflict relating to the original question, thus consequently reinforcing curiosity. At an individual level, knowledge acquisition involves finding sensible explanation and new inference for facts that do not agree with existing mental schemata ([62], [63]), and can be indexed by generation of diverse problem solving approaches [63]. It also comprises comparison with existing knowledge or search for relevant knowledge through external resources to reduce simultaneous opposing beliefs that might stem from the investigation [35]. At an interpersonal level, knowledge acquisition comprises revelation of uncertainties in front of group members [48], joint creation of new interpretations and ideas, engagement in argument to reduce dissonance among peers [64], and critical acceptance of what is told [48].

Finally, we call the third function group Intensification of Knowledge Identification and Acquisition. The intensity of curiosity, or the desire for new knowledge is influenced by factors such as the confidence required to acquire it [21], its incompatibility with existing knowledge, existence of a favorable environment [35] etc. At an individual level, intensification of knowledge identification and acquisition can stem from factors such as anticipation of knowledge discovery [38], interest in the topic [47], willingness to try out tasks beyond ability without fear of failure [65], taking ownership of own learning and being inclined to see knowledge as a product of human inquiry [48]. These factors can subsequently result in a state of increased pleasurable arousal [20]. At an interpersonal level, intensification of knowledge identification and acquisition is influenced by the willingness to get involved in group discussion and the tendency to be part of a cohesive unit in the pursuit of instrumental objectives and/or for the satisfaction of a group member’s affective needs [35]. Such willingness can span from the spectrum of merely continuing interacting to pro-actively reacting to the information others present [40]. Various interpersonal factors play out along different portions of this spectrum. Salient ones include interest in knowing more about a group member [66], promotion of an unconditional positive and non-evaluative regard towards them [38], tendency of spontaneous pickup of behavior initiated by a group member (where the initiator did not display any communicated intent of getting the others to imitate) [35], and awareness of one’s own uncertainty being shared or considered legitimate by those peers [12]. All of these factors can subsequently result in cooperative effort to overcome common blocking points for the group [38].
3.2 Behaviors Fulfilling Putative Functions of Curiosity

Behavioral episodes including language and associated multimodal communicative signals (e.g., facial expressions, gaze etc) serve as both communicative markers, i.e. they provide evidence for presence of curiosity of group members, and mind markers, i.e. they shape group members' understanding and expectation of how to approach the task, along with conceptualization and construction of the associated knowledge ([67], [68]). They can (i) contain a single action or multiple co-occurring or contingent actions made by one or more individuals, (ii) be purposeful or non-purposeful, because the underlying human strategy that governs the sequence of behaviors is unknown. Our review of prior research in psychology and learning sciences led us to link the behaviors with their functions in evoking curiosity, and organize these behaviors into four clusters. Table 1 illustrates examples of these behavior clusters from empirical observation of informal group learning activities.

Cluster 1 corresponds to behaviors that enable an individual to get exposed to and investigate physical situations, which may spur socio-cognitive processes that are beneficial to curiosity-driven learning ([20], [62]). Examples include orientation (using eye gaze, head, torso etc) and interacting with stimuli (for e.g - manipulation of objects). When looking at video segments tagged with high curiosity in our empirical data, these behaviors occur in contexts where children look at different aspects of the stimulus (e.g. - function of novel objects, physical properties of mineral samples in the science class, transition phase of dry ice samples in the summer camp etc) by orienting towards it using their gaze and torso, smelling or scratching it, rotating and trying to fit more than one object together etc. Cluster 2 corresponds to behaviors that enable an individual to actively make meaning out of observation and exploration ([20], [62], [69]). Examples include idea verbalization, justification, generating hypotheses etc. Cluster 3 corresponds to behaviors that involve joint investigation with other group members ([20], [62], [69]). Examples include arguing, evaluating problem-solving approach of a partner (positive or negative), expressing disagreement, making suggestions, sharing findings, question asking etc. In video segments tagged with high curiosity, these behaviors occur in contexts where children listen to other’s suggestion, express disagreement or challenge their perspective by pointing out loopholes, and engage in physical demonstration to make their point clear. Finally, Cluster 4 corresponds to behaviors that reveal affective states of an individual ([11], [70]) including expressions of surprise, enjoyment, confusion, uncertainty, flow and sentiment towards task.

We hypothesize that behaviors across these clusters will map onto one or more putative functions of curiosity, since there can be many different functions or reasons why a communicative behavior occurs. For example, in knowledge-based conflict in group work, attending to differing responses of others compared to one’s own may raise simultaneous opposing beliefs (knowledge identification). This awareness might in turn activate cognitive processes, wherein an individual may seek social support for their original belief by emphasizing its importance and validating others idea by providing justification, or, engaging in a process of back and forth reasoning to come to a common viewpoint (knowledge acquisition). Furthermore, this awareness may as well impact social and emotional

<table>
<thead>
<tr>
<th>Behavior Cluster</th>
<th>Empirical Observation (Example 1)</th>
<th>Empirical Observation (Example 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1,2</td>
<td>P1: Hey let’s wait I have an idea [idea verbalization]</td>
<td>P1: So the chain has to be like this [idea verbalization with iconic gesture]</td>
</tr>
<tr>
<td></td>
<td>P1: Let’s see what this is, but let me just, let me just... [proposes joint action, co-occurs with physical demonstration, initiates joint inquiry]</td>
<td>P1: How would that be? [question asking followed by orienting towards stimulus]</td>
</tr>
<tr>
<td></td>
<td>P2: I have no idea how to do this, but it’s making my brain think [positive attitude towards task]</td>
<td>P1: Well, I don’t want it to break, so I want it to be about... no, lets say half an... half an inch [causal reasoning to justify actions being taken]</td>
</tr>
<tr>
<td>Cluster 1,3</td>
<td>P1: Wait we need to raise it a bit higher [making suggestions]</td>
<td>P2: And the funnel can drop it into one of um... those things</td>
</tr>
<tr>
<td></td>
<td>P1: Maybe if we put it on. Umm... this thing, maybe... this is high enough [co-occurs with joint stimulus manipulation]</td>
<td>P1: If the funnel can drop it</td>
</tr>
<tr>
<td></td>
<td>P2: Why? W-Why do we need to make it that high? [disagreement and asking for evidence]</td>
<td>P1: Okay but then... even if it hits this, then we need what is this going to hit? [challenge]</td>
</tr>
<tr>
<td></td>
<td>P1: Here- let- just- just- just- make sure that its going to hit it [followed by physical demonstration/verification]</td>
<td>P1: Here- let- just- just- make sure that its going to hit it [followed by physical demonstration/verification]</td>
</tr>
<tr>
<td>Cluster 2,3,4</td>
<td>P1: Roll off into here and go in there [hypothesis generation]</td>
<td>P2: We could use this if we wanted [making suggestion]</td>
</tr>
<tr>
<td></td>
<td>P1: Okay, so how are we going to do that? [question asking]</td>
<td>P1: Lets figure this quickly... so we at least have this part done [preceded by expression of surprise and followed by trying to connect multiple objects to create a more complex object]</td>
</tr>
<tr>
<td></td>
<td>P2: It looks like something should hit the ball [making suggestion]</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Corpus examples of behavior sequences. P1 is the child with high curiosity (see section 5.1 to see how we annotated ground truth curiosity)

Fig. 2. Empirical observation across learning contexts. Left: in-lab RGM building; Middle: in-school STEAM class; Right: science summer camp
processes, where an individual may perceive a conflict differently and their emotions felt and expressed might vary depending on relation with and perception of the source of conflict, for e.g, is it a friend/stranger, more competent/less competent, more cooperative/less cooperative group member that raises conflict, and therefore take the next action of resolving that conflict differently (intensification of knowledge identification and acquisition). We intend to discover prominent mappings between functions described in section 3.1 and behaviors described in section 3.2 more formally in a data-driven way in section 6.

4 Data Corpus

We carried out extended empirical observation of small groups of 9-14 year old children engaged in hands-on learning activities across learning contexts (see Fig. 2). Our goal behind collecting such rich multimodal data was to be able to then annotate ground truth curiosity, along with manual or automatic detection of verbal and nonverbal behaviors that fulfill putative functions of curiosity (see section 3.2), and subsequently statistically verify predictive relationships between them (see section 6).

In order to capture the simultaneous dynamic of complex behaviors of multiple participants, we developed a novel human behavior recording apparatus to collect fine-grained video and audio footages at three different levels (Fig. 3: frontal view for automated detection of participants facial expression, head and gaze gesture, single view and group view for analyzing other individual and social behaviors. The apparatus (Fig. 4) contains (1) egocentric recording (frontal view): four Logitech C920 Pro Webcam devices; (2) exocentric recording (single and group view): eight camcorder recorders. We also used four lapel-microphones (Sennheiser ew100 G2 and Samson UHF Micro 32) to separate individuals speaking.

4.1 In-Lab Study

The in-lab study comprised forty-four participants in 5th or 6th grade (16 male and 28 female, aged between 10-12, average age 11.2), who collaboratively build a Rube Goldberg machine (RGM) for about 35-40 minutes. The RGM task included building several chain reactions that were meant to be triggered automatically to trap a ball in the cage (without external human support), and the machine was created using variety of simple objects such as rubber band, pipe cleaner, toy car and clothespin etc. We designed the RGM building activity because it supports learning of key science knowledge for students in 5th and 6th grade as defined by the Pennsylvania Department of Education [71] such as force, motion and energy transfer, and enables collaborative hands-on learning and creative problem solving [72].

Participants were recruited through recruiting flyers sent to local public schools, parent mailing-lists of the university, and advertisement on social media and in public spaces. All participants were remunerated with $30 cash. The study took place in the usability laboratory in the Human Computer Interaction institute of Carnegie Mellon University. There were 12 study sessions in total, each one including one group of three to four students and lasting for one and half hour. The procedure for one study session was as follows: (1) 10 minutes for an ice-breaking game; (2) 5 minutes free exploration of materials easily found at home; (3) 5 minutes of introducing participants to Rube Goldberg machines (RGM); (4) 30 minutes for collaboratively building the RGM; (5) a 5 minute opportunity to demonstrate the built RGM; (6) a 10 minute interview, and optionally (7) 5 minutes free exploration of a pre-made RGM; (8) 20 minutes to enjoy snacks and dispense remuneration.

4.2 In-School Study

Twenty participants in 6th grade participated in the in-school study. Participants were recruited by teachers in a
local public school. There was no remuneration. The study took place in the 6th grade science class and STEAM class for two weeks. There were in total eight observation sessions, four sessions for science class and four sessions for STEAM class, with each session lasting about 45 minutes. The main activities of the science class were earth science learning, and the main activities of the STEAM class were robotics, programming, and crafting. Both classes were divided into small groups with 3-4 students.

There were several constraints for data collection in a classroom, including large-scale observation (3-4 groups of students at the same time), noisy environments, and confined space and time for set up (minimizing obstacles for teachers and teaching assistants to move around during group discussion, and quick equipment setup during the short breaks between two classes). To accommodate the above constraints, we used a lightweight recording setup that includes two camcorder recorders to capture students from the opposite side of the table, and one lapel microphone in the middle of the table. The researchers also took field notes to describe student’s interpersonal learning behaviors relating to curiosity.

4.3 Science Summer Camp Study

Thirty-one participants aged 9-14 participated in the science summer camp study. Participants were recruited by teachers in a local child maker-space. We provided each participant a t-shirt as a gift. The study took place in two science summer camp sessions hosted in a local child maker-space, with each summer camp lasting one week. We observed about 30 hours in total. There was a wide variation of hands-on activities in the summer camp, such as physics, chemistry, biology, life science, robotics, and crafting. The class was divided into four to five groups.

In addition to similar constraints as the classroom, the observation at the summer camp was longer and the activities were more diverse and less structured. Therefore the layout of the class changed frequently and the classroom was noisier. We used one camcorder recorder and one lapel microphone fixed on the table for each group of children. Similar as the in-school study, the researchers took field note to describe peer-peer interaction associated with curiosity.

5 Quantitative Analyses of In-Lab Study

We now describe fine-grained quantitative analyses from a convenience sample of the first 30 minutes (out of 35-40 minutes given each group), of the RGM task (lab study) for half of the sample; that is, 22 children across 6 groups. Table 2 provides a summary of all coding metrics used in this paper. Our goal is to empirically verify the theoretical framework of curiosity proposed in section 3.

5.1 Assessment of Ground Truth Curiosity

Person perception research has demonstrated that judgments of others based on brief exposure to their behaviors is an accurate assessment of interpersonal dynamics [73]. We used Amazon’s Mechanical Turk (MTurk) platform to obtain ground truth for curiosity via such a thin-slice approach, using the definition “curiosity is a strong desire to learn or know more about something or someone”, and a rating scale comprising 0 (not curious), 1 (curious) and 2 (extremely curious). The use of crowdsourcing platforms provides benefits of a diverse sample of raters who can be accessed quickly, easily and for relatively little cost [74]. Our previous research has successfully deployed thin-slice coding for other social phenomena such as interpersonal rapport in peer tutoring using MTurk ([60], [75], [76]).

In the current study, four naive raters annotated every 10 second slice of videos of the interaction for each child presented to them in randomized order. To post-process the ratings for use, we removed those raters who used less than 1.5 standard deviation time compared to the mean time taken for all rating units (HITs). We then computed a single measure of Intraclass correlation coefficient (ICC) for each possible subset of raters for a particular HIT, and then picked ratings from the rater subset that had the best reliability for further processing. Finally, inverse-based bias correction [77] was used to account for label overuse and underuse, and to pick one single rating of curiosity for each 10 second thin-slice. The average ICC of 0.46 aligns with reliability of curiosity in prior work ([78], [79]).

5.2 Assessment of Verbal Behaviors

We adopted a mix of semi-automatic and manual annotation procedures to code 11 verbal behaviors, in line with the curiosity-related behavioral set described in section 3.2. These verbal behaviors span propositional [80] and interpersonal [81] functions of contributions to a conversation. Propositional functions are those that are fulfilled by contributing informational content to the dialog (e.g. - idea verbalization, justification etc), and interpersonal functions are those that are fulfilled by managing the relationship between the interlocutors (e.g. - social question asking, positive evaluation etc). Five verbal behaviors were coded using a semi-automatic approach - uncertainty, argument, justification/suggestion at the clause level, and agreement at the turn level. A clause contains a subject (a noun or pronoun) and a predicate (conjugated verb that says something about what the subject is or does). During a full turn, a speaker holds the floor and expresses one or more interpretable clauses.

First, a particular variant of neural language models called paragraph vector or doc2vec [82] was used to learn distributed representations for a clause/turn. This means that for every clause/turn in our data corpus, we transformed the sequence of words in it to a tuple (or vector) of continuous-valued features that characterize the semantic meaning of those words. Such feature representation implies that sentences in a test set that functionally similar to sentences in a training set can still achieve good predictions. The motivation for this approach stems from the following reasons: (i) lack of available corpora of verbal behaviors that are large enough, and collected in similar settings as ours (groups of children engaged in open-ended scientific inquiry), and hence (ii) limited applicability of traditional n-gram based machine learning models to cross-domain settings, which would result in a very high-dimensional representation with poor semantic generalization, (iii) limitations of other popular neural language models such as word2vec that do not explicitly represent word order and surrounding context in the semantic representation (in contrast, doc2vec models contain an additional paragraph token that acts as
TABLE 2
A summary of coding methods used for the annotation (detailed coding scheme present at http://tinyurl.com/codingschemecuriosity)

<table>
<thead>
<tr>
<th>Construct</th>
<th>Definition used to code/define the construct</th>
<th>Coding method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curiosity</td>
<td>A strong desire to learn or know more about something or someone.</td>
<td>Four MTurk raters annotated each 10–30 item slice; average 0.00–0.61; used inverse-based bias correction to pick the final rating.</td>
</tr>
</tbody>
</table>

**Verbal Behavior**

1. **Uncertainty**
   - Lack of certainty about one’s choices or beliefs, and is verbally expressed by language that creates an impression that something important has been said, but what is communicated is vague, misleading, evasive or ambiguous.
   - *e.g. — “well maybe we should use rubberbands on the foam pieces,” “wait do we need this thing to function it will go straight into the trap”*
   - Used a semi-automated annotation approach: after automatic labeling of these verbal behaviors, two trained raters (Krippendorff’s alpha >0.86) independently corrected machine-annotated labels; average percentage of machine annotation that remained the same after human correction was 85.9 (SD=12.71).

2. **Agreement**
   - A coherent series of reasons, statements, or facts intended to support or establish a point of view.
   - *e.g. — “you could kick a ball to kick something,” “you are adding more weight there which would make it fall down”*
   - Used a semiautomated annotation procedure due to unavailability of existing training corpus (Krippendorff’s alpha >0.78 between two raters).

3. **Justification**
   - The action of showing something to be right or reasonable by making it clear.
   - *e.g. “I think we need more weight to like push it down,” “You might want to like make it fall down either” (Quote) — “Yeah we need more weight on this side” (Response).

4. **Suggestion**
   - An idea or plan put forward for consideration.
   - *e.g. “you have to kick a ball to kick something,” “you are adding more weight there which would make it fall down”*

5. **Agreement**
   - Harmony or accordance in opinion or feeling; a position or result of agreeing.
   - *e.g. “You should move up here to like make it more manageable”*

6. **Question Asking**
   - Asking any kind of questions related to the task or non-task relevant aspects of the social interaction.
   - *e.g. “Do you want to build something like a chain reaction or something like that?”, “do you think we need to make it that high?”*

7. **Idea Ventilation**
   - Explicitly saying out an idea, which can be just suggested by an individual but requires something that results in or other action.
   - *e.g. “It should now be about 250”*

8. **Sharing Findings**
   - Explicitly sharing information or communicating results, findings and discoveries to other group members during any stage of the scientific inquiry process.
   - *e.g. “look how I'm gonna see I'm gonna trap it,” “look I made my pillar perfect”*

9. **Hypothesis Generation**
   - Expressing one of more different possibilities or theories to explain a phenomenon by giving reasons between two or more variables.
   - *e.g. “We could use scissors to cut off the baby's head which would cause enough friction”, “we should see if we could make it solid so that the force of hitting it makes it fall”*

10. **Task Sentiment (Positive/Negative)**
    - A view or attitude (emotional valence) toward a situation or event, an overall opinion towards a subject matter.
    - *e.g. “oh it's the coolest cage I've ever seen,” “I'm gonna try to find out at the claw for the end to make this game go and fall”, “I'm not seeing anything direct at this cage,” “but I don't know here to make it better”*

11. **Evaluation (Positive/Negative)**
    - Characterization of how a person assesses a previous speaker’s action and problem-solving approach. It can be positive or negative.
    - *e.g. “oh that's a pretty good idea - that was a good idea,” “let's make this thing elevated and make it go down,” “oh wait don't! you're not making anything ever here,” “no it won't like that otherwise it will be stuck”*

**Non-verbal Behavior**

- **Facial**
  - **AU 3 (cheek raiser)**
  - **AU 23 (lip tightener)**
  - **AU 6 (raised lower eyelid)**
  - **AU 5 (upper lid raise)**

- **Task-Related**
  - **AU 7 (lid tightener)** and **AU 12 (lip corner puller)** and **AU 25 (lips part)**.

- **Silence**
  - The amount of time spent when holding floor before allowing a response.

**Turn Taking**

- ** aggregate**
  - A weighted product of number of group members whose turn was responded to (activity) and total time that other people spent on their turn (time handing over the floor).

- **sentential**
  - A weighted product of number of group members to whom floor was given to (sentences).
machine annotated labels that did not change even after the human de-noising step was 85.9 (SD=12.71), meaning majority of labels were correctly predicted in the first place. This was also reflected in a good cross validation training performance of the models (right column of Table 3). Six other verbal behaviors (question asking (on-task, social) (α=1), idea verbalization (α=0.761), sharing findings (α=0.3), hypothesis generation (α=0.79), attitude towards task (positive, negative) (α=0.835), evaluation sentiment (positive, negative) (α=0.784)) were coded using a traditional manual annotation procedure due to unavailability of existing training corpus.

5.3 Assessment of Nonverbal Behaviors

A rich body of existing work in affective computing for learning [89] has comprehensively described the study of emotions along 3 main sub-components: (i) subjective sensation (awareness of the emotion), (ii) physiological manifestations (heart rate etc), (iii) observable behavior manifestations (activation of certain facial landmarks etc). In our current work, we focused on facial landmarks as the starting point. Our motivation for coding nonverbal behaviors in the context of studying curiosity is inspired by prior theoretical and empirical research, which has identified the facial action units accompanying the experience of certain emotions that often co-occur with curiosity [78], and has discovered consistent associations (correlations as well as predictions) between particular facial configurations and human emotional or mental states [70], [78], [90].

We used automated visual analysis to construct five feature groups corresponding to emotional expressions that provide evidence for presence of the affective states of joy, delight, surprise, confusion and flow (a state of engagement with a task such that concentration is intense). A simple rule-based approach was followed (see Table 2) to combine emotion-related facial landmarks, which were previously extracted on a frame by frame basis using a state-of-the-art open-source software OpenFace [91]. We then selected the most dominant (frequently occurring) emotional expression for every 10 second slice of the interaction for each group member, among all the frames in that time interval.

Automated visual analysis was also used to capture variability in head angles for each child in the group, which correspond to head nods (i.e. pitch), head turns (i.e. yaw), and lateral head inclinations (i.e. roll). The motivation for using head movement in our curiosity framework is inspired by prior work in the multimodal analytics [92], [93] that has emphasized contribution of nonverbal cues in inferring behavioral constructs such as interest and involvement that are closely related to the construct of curiosity. By using OpenFace [91], we first performed frame by frame extraction of head orientation, and then calculated the variance post-hoc to capture intensity in head motions for every 10 second of the interaction for each group member. Since head pose estimation takes as input facial landmark detection, we only considered those frames that had a face tracked and facial landmarks detected with confidence greater than 80%.

5.4 Assessment of Turn Taking Dynamics

While the annotated verbal behaviors fulfill propositional and interpersonal conversational goals in the social interaction, the interactional function of contributions to a conversation is captured by turn-taking behaviors. Interactional discourse functions are “responsible for creating and maintaining an open channel of communication between the participants” [80]. The motivation for capturing turn taking in the current research stems furthermore from prior literature that has used measures such as participation equality and turn taking freedom as indicators of involvement in small-group interaction [94].

Specifically, in the current work, we designed two novel metrics using a simple application of social network analysis for weighted data. By representing speakers as nodes and time between adjacent speaker turns as edges, the following two features were computed for each group member for every 10 seconds: (i) TurnTakingIndegree, which was a weighted product of number of group members whose turn was responded to (activity) and total time that other people spent on their turn before handing over the floor (silence), and was quantified as activity$^{1-\alpha}$ silence$^{\alpha}$. Since high involvement is likely to be indexed by higher activity and lower silence, $\alpha$ was set to -0.5, (ii) TurnTakingOutdegree, which was a weighted product of number of group members to whom floor was given to (participation equality), and the amount of time spent when holding floor before allowing a response (talkativeness), and was quantified by participation equality$^{1-\alpha}$ talkativeness$^{\alpha}$. Since higher participation equality and talkativeness are favorable, $\alpha$ was set to +0.5. We will use these two metrics in our empirical validation of the theoretical framework of curiosity.

6 Empirical Validation of the Proposed Theoretical Framework of Curiosity

We used a “multiple-group” version of continuous time structural equation model (CTSEM) [95] to evaluate the proposed theoretical framework of curiosity, and statistically verify the predictive relationships between ground truth curiosity (that we formalized as our manifest variable), functions described in our theoretical framework (that we

<table>
<thead>
<tr>
<th>TABLE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Verbal Behavior</strong> (Krippendorff’s α for human judgment)</td>
</tr>
<tr>
<td>1. Uncertainty [0.78]</td>
</tr>
<tr>
<td>2. Argument [0.792]</td>
</tr>
<tr>
<td>3. Justification (process [0.638], causal [0.710], model [0.828], example [0.718])</td>
</tr>
<tr>
<td>4. Suggestion [0.608]</td>
</tr>
<tr>
<td>5. Agreement [0.835]</td>
</tr>
</tbody>
</table>
formalized as latent variables) and multimodal behaviors (that we formalized as time-dependent predictors). We first present a formal description of the technical background underlying our latent variable analyses in section 6.1, and then describe our approach in more detail in section 6.2, concluding this section with model results in section 6.3.

6.1 Technical Background of CTSEM

Structural Equation Model (SEM) is a statistical technique for testing and estimating causal relationships using a combination of statistical data and qualitative causal assumptions. Conventional SEM procedure assumes independent observations and thus cannot be applied directly to analyze auto correlated time series data arising in multimodal human behavior analysis. This points towards consideration of a Dynamic Bayesian Network like model that can explicitly model temporal dependencies between the latent random variables across time-steps. At present, applications of such dynamic models in the social and behavioral sciences are almost exclusively limited to discrete time models, where it is assumed that time progresses in discrete steps, and that time intervals between measurement occasions are equal. In many cases, these assumptions are not met, resulting in biased parameter estimates and a misunderstanding of the strength and time course of effects. Continuous time SEM models overcome these limitations by using multivariate stochastic differential equations to estimate an underlying continuous process and recover underlying latent or hidden causes linking entire sequence.

Formally, a multivariate stochastic differential equation for a latent process of interest in CTSEM can be written as

\[
\frac{d\eta_i(t)}{dt} = A\eta_i(t) + Bz_i + M\chi_i(t) + GdW_i(t) + \xi_i \quad \text{(Structural part of the SEM model)},
\]

where \( A \) is the drift matrix that models auto effects on the latent variable has on itself on the diagonals, and cross effects to other latent processes on the off-diagonals, in turn characterizing temporal relationships between the processes. \( \xi_i \) determines the long-term level of the latent process. Matrix \( B \) and \( M \) represent the effect of time independent and time dependent variables on the latent process. Time independent predictors would typically be variables that differ between subjects, but are constant within subjects for the time range in question (for e.g - a trait curiosity questionnaire collected at the beginning). On the other hand, time dependent predictors vary over time and are independent of fluctuations in of the latent processes in the system, and can be treated as a simple impulse form, in which the predictors are treated as impacting the processes only at the instant of an observation. Matrix \( G \) represents the effect of noise or the stochastic error term \( dW_i(t) \) on the change in the latent process. \( Q = G*G^T \) represents the variance-covariance matrix of diffusion process in continuous time. The essence of diffusion processes is to capture very slow patterns of change in latent variable, upwards or downwards trajectories that are maintained over many observations (persistence) depending on contextual circumstances. Furthermore, this latent process can be used to predict manifest variables of interest using the equation

\[
y_i(t) = \Lambda\eta_i(t) + \zeta_i(t) \quad \text{(Measurement part of the SEM model)},
\]

where \( \Lambda \) is a matrix of factor loadings between the latent and manifest variables and \( \zeta_i \) is the residual (error) vector.

A Kalman filter can be used to fit CTSEM to the data and obtain standardized estimates for the influence of behaviors on latent functions, and in turn these latent functions on curiosity. From a theoretical standpoint, Kalman filter is an algorithm permitting exact inference in a linear dynamical systems. It uses a series of measurements observed over time (containing statistical noise and other inaccuracies) to produce estimates of unknown variables that tend to be more precise than those based on a single measurement alone. It is a state space model described by a (i) state equation that describes how the latent states change over time and is analogous to structural part of the SEM model, and (ii) output equation that describes how the latent states relate to the observed states at a single point in time (how the observed output is produced by the latent states), and is analogous to measurement part of the SEM model.

Finally, an important point to note is that when there are multiple groups in a dataset (for e.g, we have 6 groups in our corpus), a “multiple group” version of CTSEM should be used. It allows investigation of group level differences and helps understand variability in model parameters across different groups.

6.2 Application of CTSEM to In-Lab Study Corpus

Since knowledge identification and acquisition are closely intertwined with knowledge seeking behaviors and it is hard to draw a distinction between these putative underlying mechanisms based on observable or inferred multimodal behaviors, we formalized them under the same latent variable. The final set of latent functions for our theoretical framework that we statistically verified therefore included: (i) individual knowledge identification and acquisition, (ii) interpersonal knowledge identification and acquisition, (iii) individual intensification of knowledge identification and acquisition, (iv) interpersonal intensification of knowledge identification and acquisition.

Two versions of CTSEM were run. In first version, we specified a model where only factor loadings between the manifest variable and latent variables in measurement part of the model were estimated for each group distinctly (we report the average and standard deviation across the 6 groups in Fig. 5), but all other model parameters including those belonging to structural part of the model were constrained to equality across all groups (Model\textsubscript{constrained}) and then estimated freely. This means that matrices \( A, B, M, G \) and \( A \) were freely estimated. Since the form of a behavior does not uniquely determine its function, nor vice-versa, we did not pre-specify the exact pattern of relationships between behaviors and functions to look for/estimate. In second version of the model, all parameters for all groups were estimated distinctly (Model\textsubscript{free}).

Technically, the rich representational capacity of “multiple-group” CTSEM allows running these two separate models. However, analytically, the decision to separately run these two models was based on the intuition that while the relationships between appearance of behaviors and their contribution to the latent functions of curiosity would remain the same across groups, the relative contribution of interpersonal or individual tendencies for knowledge identification, acquisition and intensification would vary based on learning dispositions of people towards seeking
the unknown. This intuition stemmed from prior literature in learning analytics that has looked into measuring learning dispositions [48], an important dimension of which is the ability of learners to balance between being sociable and being private in their learning, i.e. not being completely independent or dependent, but rather working interdependently. We hypothesized that this dimension will impact curiosity differently, especially since our data comprised a group learning context, and therefore expected Model constrained to fit the data better than Model free.

6.3 Model Results and Discussion

An empirical validation confirmed our hypothesis of Model constrained fitting to the data better than Model free. The Akaike Information Criterion (AIC) for Model constrained (933.48) was \(\sim 3\times\) lower than Model free (2278.689). We now illustrate results of the CTSEM (Model constrained) in Fig. 5, depicting links with top ranked standardized estimates between behaviors and latent variables. In few cases, we also added links with the second highest standardized estimate if they clarified our interpretation of the latent function.

Overall, these results provide confirmation of correctness of the theoretical framework of curiosity along three main aspects: (i) The grouping of behaviors under each latent function and their contribution to individual and interpersonal aspects of knowledge identification, acquisition and intensification aligns with prior literature on the intrapersonal origins of curiosity, but also teases apart the underlying interpersonal mechanisms, (ii) There exists strong and positive predictive relationships between these latent variables and thin-slice curiosity, (iii) Knowledge identification and acquisition have stronger influence to curiosity than knowledge intensification, and interpersonal-level functions have stronger influence compared to individual-level functions. We now discuss latent functions and associated behaviors, ordered by the degree of positive influence on curiosity.

First, Interpersonal Knowledge Identification and Acquisition shows the strongest influence to curiosity among the four latent functions (2.612 ± 0.124). The natural merging of knowledge identification and knowledge acquisition corroborates with the notation that one person’s knowledge seeking may draw attention of another group member to a related knowledge gap and escalate collaborative knowledge seeking. Behaviors that positively contribute to this function are mainly from cluster 3 (sharing findings, task related question asking, argument, and evaluation of other’s idea). In addition, nonverbal behaviors including head turn and turn taking dynamics (indegree) are also related to this function, which support the idea that higher degree of group members’ interest and involvement in the social interaction stimulates awareness of peer’s ideas, subsequently leading to knowledge-seeking via social means in order to gain knowledge from the experience of others and add that onto one’s own direct experiences.

Second, Individual Knowledge Identification and Acquisition shows a strong influence to curiosity (2.149 ± 0.066). Similar to the interpersonal level function, knowledge identification and acquisition merge into one coherent function, as knowledge-seeking behaviors can spark new unknown or conflicting information within the same individual. Behaviors from cluster 2 (hypothesis generation, justification, idea verbalization) and cluster 4 (confusion, joy, surprise, uncertain, positive sentiment towards task) mainly contribute to this function. Head nod, as indicative of positive feelings towards the stimulus due to its compatibility with the response [96], maps to this function as well. Finally,
we find that turn taking (indegree and outdegree) and social question asking contribute positively to individual knowledge identification and acquisition. Interest in other people reflects a general level of trait curiosity and influences inquisitive behavior [66].

Third, we find that a relatively small group of behaviors including agreement, idea verbalization and lateral head inclination have predictive influence on the latent function of Interpersonal Knowledge Intensification, which in turn has a high positive influence on curiosity (1.756 ± 0.238). Agreement may contribute to information seeking by promoting acceptance and cohesion. Working in social contexts broadcasts idea verbalization done by an individual to other group members, which might in turn increase their willingness to get involved. Lateral head inclination during the RGM activity is associated with intensive investigation of the RGM solution offered by both oneself and other group members. Overall, engagement in cooperative effort to overcome common blocking points in the group work may result in intensifying knowledge seeking.

Finally, the latent function of Individual Knowledge Intensification has the least comparative influence on curiosity. It is associated with non-verbal behaviors such as head nod and emotional expressions of positive affect (flow, joy and delight), which function towards increasing pleasurable arousal. In addition, surprise and suggestion also positively influence this latent function, and signal an increased anticipation to discover novelty, conceptual conflict, and correctness of one’s own idea. Interestingly, results also show that negative sentiment about the task positively influences an individual’s knowledge seeking behaviors. A qualitative examination of the corpus reveals that such verbal expressions often co-occur with evaluation made by a group member within the same 10 second thin-slice that signals a desire for cooperation. Thus, a potential explanation of this association is that expressing negative sentiment about task may signal hardship, which draws group member’s attention and increases chances of receiving assistance, thus increasing engagement in knowledge seeking.

7 Implications for Learning Technologies

While the importance of metacognitive factors in learning ([49], [97], [98]) has long been recognized, only few current learning technologies actually aim to foster them (for e.g, see [99], [100], [101], [102], [103]). Reasons include (i) lack of theoretical formalisms and real-time measurement approaches to capture the intricate nature of metacognitive and socioemotional factors such as creativity, curiosity, grit, help-seeking, self-explanation ability etc [104], and (ii) dearth of an operational way to embed this theoretical understanding into computational models that can leverage mapping between behaviors and their putative underlying mechanisms to offer scaffolding [105]. The research presented in this work therefore goes beyond prior work that has worked on inferring curiosity directly from visual and vocal cues (for e.g, [78], [79], [106]), without adequate consideration of underlying mechanisms that link these low-level cues to curiosity, as well how these cues interact with group dynamic behaviors and other verbal cues.

Knowing what forms of verbal and nonverbal behaviors and their corresponding functions are good indicators of curiosity in the human-human interaction allows us to better understand and design teaching tools and technological learning environments that can specifically and intentionally look for opportunities to use strategies (that is, to be tactical) to scaffold, maintain and evoke curiosity. Concretely, this challenging undertaking includes development of (i) a curiosity perception module that accurately senses a comprehensive subset of verbal and non-verbal behaviors by using available sensors (for e.g, cameras, microphones and other biometric devices) in real-time interaction, (ii) a curiosity reasoner that outlines how a learning technology could choose to act a certain way so as to support the same functions. We believe that dissociating different dimensions along which scaffolding can be provided holds the key to combating lack of curiosity in small group work. These implications go beyond prior work on pedagogical agents/robots ([52], [53]) by not making assumptions regarding what observable behaviors can serve as proxies for lack of learner’s curiosity. Instead, we take a step towards reasoning and understanding the presence or absence of such curiosity by introducing the theoretically grounded layer of functions that can be fulfilled (by behaviors).

In the two subsections below, we now further expand on how these functions can be supported (using strategies and associated tactics). The motivation for finding strategies and particular instantiations of these strategies (tactics) to foster curiosity comes from both prior literature [107] and our own empirical data observation, where we find that (i) a curious child may not always make an explicit attempt to raise curiosity of another group member (perhaps because it needs increased cognitive/social effort), and (ii) disinterested children may be increasingly cut-off from the core interactions of the group over time, and therefore it’s important to find ways to reignite their interest.

7.1 Strategies for Supporting Functions of Curiosity

We define strategies as intentional decisions made by a third party (learning technology or human coach/peer) in the service of facilitating curiosity. This means that strategies serve as vehicles for making an influence attempt in the group, and directly (for e.g - affecting gains/costs) or indirectly (for e.g - controlling critical environmental aspects) affecting the behavior of a group member [35]. We believe that strategies should support underlying functions that contribute to curiosity, since this enables computer support to address the root cause of an undesirable behavior, or provide a reinforcing means for the root cause of a desirable behavior. The strategies that we are going to describe below can support interaction regulation to various extents. For learning technology, this implies that when administering support, success of a strategy usage can be determined broadly by the extent to which it increases thin-slice curiosity in subsequent time-intervals, and specifically by the probability that it actually leads to expected behavior(s) by a target child. We propose a categorization of strategies into four clusters in alignment with the functions of curiosity.

The first cluster of strategies facilitate the knowledge identification function that raises awareness of knowledge to be acquired. Overall, they help clarify issues in the group to help members realize potential new knowledge to seek,
and help save face by providing a graceful means of accepting compromises ([35], [108]). Representative examples from the literature and data include: (i) making group members aware of conflict between approaches, (ii) raising conflict and, (iii) raising awareness of novel or complex stimulus.

The second cluster of strategies facilitate the knowledge acquisition function that enhances information seeking skills. Overall, they stimulate critical thinking by helping group members develop new interpretations and consider alternative perspectives [109], and prevent them from valuing their cohesiveness and relationships with others so much that they avoid conflict and challenging each other’s ideas [35]. Representative examples include: (i) inviting group member’s thinking in terms of them providing multiple answers that require integration and reasoning, (ii) inviting thinking in non-routine ways, (iii) nurturing thinking challenge by provoking group members out of their comfort zone and encouraging them to rethink/defend their responses, (iv) helping group members find causal relationships between processes, (v) encouraging collaborative reasoning, (vi) mediating conflict (assisting reconciliation by providing constructive responses to conflict).

The third cluster of strategies provide a supportive environment to facilitate the intensification of knowledge identification and acquisition function. Overall, they help create a friendly climate and honor knowledge gained through trial and error, promoting an idea of learning as a proactive process that may involve failure. They also assist group members in maintaining productive framing of the interaction, or shifting to productive framing, where it is appropriate to explain or grapple with own intuitions [68]. This helps group members see risk as a positive social value towards task completion, and contributes towards building a psychologically safe environment [39], where children can open up to group members with reduced uncertainty of acceptance or perception of incompetency. In addition, an infectious winning attitude is radiated that instills a similar willingness to pursue knowledge ([35], [108]). Representative examples include: (i) maintaining enthusiastic attitude in the group, (ii) encouraging and modeling exploratory behavior, (iii) promoting risk-taking, (iv) avoiding threat, personal judgment, harsh evaluation and criticism, and (v) being warm, accepting and supportive.

Finally, the fourth cluster of strategies serve to sustain attention and interest to facilitate the intensification of knowledge identification and acquisition function. Overall, they capture group member’s interest by raising awareness of sources of novelty and uncertainty in the environment and stimulating an attitude of inquiry by provoking them to find sensible explanations for facts that do not agree with existing mental schema ([47], [63]). Representative examples include: (i) perceptual arousal, (ii) inquiry arousal, and (iii) variability.

7.2 Tactics for Exercising a Particular Strategy

We define tactics as particular discrete and observable ways in which a strategy can be exercised. The rationale behind usage of tactics is that certain forms of group interaction are more effective for raising curiosity than others, and that it is worthwhile to make some arrangements for the appearance of such interactions as they may not occur naturally - for instance sharing findings with peers, showing interest in their work, engaging in jointly conducting inquiry with them etc. This rationale is inspired from the traditional computer-supported collaborative learning literature on scripting [110], however, we propose a more real-time version of such scripting-based approaches. Our interest is rather in regulation of the small group interaction “on the fly” [111] by continuously comparing the current curiosity level with a target configuration (for e.g - if the likelihood of being in high curiosity level in the following time intervals exceeds certain threshold), and exercising tactics to restore equilibrium whenever there is a discrepancy in the current and target curiosity level. Drawing parallels with the scripting literature, we posit that tactics for providing social scaffolding for curiosity are equivalent to induced micro-scripts [110] that are embedded in a game-play environment to support awareness and coordination of group activities. We now describe these tactics falling into each of the four strategy clusters (as described in section 7.1).

Tactics that fall under the strategy of facilitating knowledge identification include: (i) providing contrasting cases, (ii) informing knowledge awareness of a group member having same or different knowledge, (iii) creating paradoxes, (iv) helping group members notice novel features of the stimulus.

Tactics that fall under the strategy of facilitating knowledge acquisition include: (i) open-ended question asking (e.g - How might you solve this problem?), (ii) encouraging multiple responses to an initiating question (e.g - What do you think about what child X did?), (iii) challenging group member’s responses (e.g - I don’t think this will be really sturdy though), (iv) asking group members to make an explicit link between ideas, representations and solution strategies (e.g - what’s your evidence for that?), (v) making group members take positions on a big question raised by a task issue and then present reasons and evidence for and against, (vi) encouraging use of connective words (because, so, if, then etc) and performative verb phrases (i think, i know), (vii) encouraging question-asking, (viii) providing materials and space to help group members create new hypotheses, (ix) fostering open dialogue in the group by facilitating the communication process (e.g - Person A, can you explain what person B said?), and making sure group members are aware of their own roles and general goals of the group (e.g - what are our priorities?), (x) providing feedback on reasoning and proposing remedial action, (xi) regulating turn taking so that both parties have a chance to express what happened, (xii) encouraging group members who are in conflict to paraphrase each other’s position, and (xiii) offering proposals for alternative solutions.

Tactics for provision of supportive environment that fall under the strategy of intensifying knowledge identification or acquisition include: (i) expressing excitement about the task, (ii) expressing curiosity-orientation behavior (e.g - looking at stimulus with surprise, expressing interest in individuals and activities using gaze or body orientation), (iii) rewarding risk-taking, (iv) encouraging low-risk self-disclosures to facilitate social information seeking, (v) providing encouragement and positive feedback for effort, (vi) expressing empathy for individual emotion, value and needs, (vii) rewarding social interaction, and (viii) deliber-
ately assigning a dominant or central role to shy or disinter-
ested group member.

Finally, some tactics for sustaining attention and interest
that fall under the strategy of **intensifying knowledge iden-
tification or acquisition** include: (i) creating wonderment
by unexpected environmental changes (e.g. - using light
intensity to emphasize a stimulus), (ii) injecting emotional
material into the interaction to emphasize meaningfulness
of the activity/task and group member’s effort, (iii) helping
group members notice novel features of the stimulus, (iv)
progressive fact disclosure, (v) bringing variations in inter-
action styles, (vi) presenting concrete analogies.

### 8 Limitations and Future Directions

The current analyses are part of a larger research effort to
understand and implement the social scaffolding of curios-
itv through an ECA [14]. The theoretical framework
presented in this paper lays foundation of a computational
model of curiosity that can enable an ECA to sense real-time
curiosity level of each member in small group interaction.
Towards this end, we have already begun initial investiga-
tion of fine-grained sequential patterns of observable behav-
iors that impact own or other group members curiosity in
human-human interaction [112].

However, it is important to note that despite existence
of learning opportunities, student’s unwillingness to learn
and explore can stem from multiple sources, such as un-
awareness of new information that is to be learned, lack
of knowledge (competency) and information-seeking skills,
lack of support from the environment, shyness in talk-
ing to group members, being perceived as incompetent
or uncertainty about being accepted. Therefore, we must
acknowledge that although we outlined individual and
interpersonal functions of curiosity in detail, it is still hard
to quantitatively determine the cause of lack of curiosity
in a fully automated manner. Future work could examine
usage of interactive questioning techniques to complement
stealth measurement approaches described in this paper for
inferring the causes of high or low curiosity. The prospect
of developing mentalizing capabilities (via perception of
social cues to form a predictive model of group members)
[113] also holds immense promise in reasoning about causal
mechanisms underlying human behaviors.

In the current work, we focused more on the epistemic,
rather than the perceptual dimension of curiosity [20].
We now intend to broaden our observable manifestations
of curiosity by studying risk-taking in open-ended group
work. Prior theoretical frameworks [114] posit that such
risk-taking behavior in play consists of contrasting arousal-
avoing and arousal-seeking states, which are strongly re-
lated to the perceptual dimension of curiosity. Furthermore,
the current work did not consider manipulative actions
on artifacts embedded in the learning environment, and
other forms of physical demonstration in the hands-on task.
These hand movements signal an individual’s investment
into seeking new knowledge by gathering information [115],
which in turn shapes their curiosity. We have therefore
started annotating hand gestures (both manipulative actions
and communicative gestures) including epistemic hand ac-
tions, symbolic (metaphoric, iconic) and deictic gestures, to
investigate their co-occurrence with verbal behaviors.

Also, since our empirical data does not always provide
evidence for naturally occurring curiosity-related behaviors,
the effectiveness of specific strategies/tactics as described
in the implications is yet to be confirmed through human
coach/peer or computer supported situations. To this end,
we have started developing a prototype of an ECA in the
form of an intelligent virtual peer that aims to elicit cu-
riosity for young children in a collaborative tabletop game.
We applied heuristics [112] derived from our theoretical
framework (described here, and in [116]) in designing social
interactions for the virtual peer to fulfill key curiosity drives
that support children’s cognitive and social engagement in
identifying and acquiring desired knowledge. The intelli-
gent virtual child will serve as the apparatus in a Wizard-
of-Oz experimental study that aims to further validate our
theoretical framework through analysis of child-agent inter-
action. This validation, in turn, may identify future research
questions and hypotheses that lead to the refinement and
iteration of the theoretical framework of curiosity.

In conducting empirical evaluation to verify the effec-
tiveness of supporting curiosity in small group learn-
ing, some challenges include (i) combating inaccuracy in
detecting multimodal behavioral episodes and assembling
different multimodal perceptions from group members for
curiosity inference, (ii) developing planning algorithms for
selecting responding strategies based on current percepts
and tying together associated sets of tactics, (iii) using
appropriate evaluation criteria to test effectiveness of the
selected strategy on subsequent human behavior.

Finally, we must acknowledge some methodological lim-
itations inherent in this work. First, the small sample size
warrants attention to generalizability of these findings to
other STEM and non-STEM disciplines of learning. Second,
the reliability of thin-slice annotation of ground truth (cu-
riosity, in our case) via crowdsourcing platforms can be
improved by varying more carefully factors such as the
time-scale (granularity) of ratings, rating scale, task setup
on the crowdsourcing platform etc. Third, our approach
of combining machine annotation with human judgment
for annotation of verbal behaviors (section 5.2) favors re-
producibility, speed and scalability, without compromising
on inter-rater reliability. Despite this rigor, going through
machine annotated labels and evaluating their accuracy
(de-noising process) is cognitively a different task than if
those labels were not there in the first place (meaning that
a completely manual annotation approach had been fol-
lowed). Future work could have some intermediate points
during this de-noising process, where the initial inter-rater
reliability for human judgment (as reported in left column
of table 3) could be re-evaluated to ensure its consistency.
Fourth, latent variable models that we used for empirically
validating our proposed theoretical framework of curiosity
(section 5) are limited by their ability to make causal in-
ferrances, especially in cross-sectional datasets [117]. Fifth,
we used a crude proxy to infer emotional states from facial
landmarks in this paper, and future work could adopt com-
plementary predictive modeling approaches [118]. While fa-
cial expressions have the advantage of being observable and
being detected using current computer vision approaches
with high accuracy, we acknowledge that they can often be
polysemous, ambiguous, and be voluntarily camouflaged
for social reasons, and these subtle distinctions between underlying mechanisms cannot be teased apart easily.

9 Conclusion

In this work, we articulated key social factors that appear to account for curiosity in learning in social contexts, proposed and empirically validated a novel theoretical framework that disentangles individual and interpersonal functions linked to curiosity and behaviors that fulfill these functions. We found strong positive predictive relationships of the interpersonal functions of knowledge identification, acquisition and intensification on curiosity, which reinforces our original hypotheses about the social nature of curiosity and the need to disentangle its interpersonal precursors from its individual precursors. Through the design of learning technologies and confirming their effectiveness, we hope to provide additional pedagogical instructions for school teachers to help children with diverse socio-economical background develop knowledge-seeking skills driven by intrinsic curiosity, support each other during scientific inquiry, and obtain equal opportunity to fulfill scientific citizenship.

Acknowledgments

We would like to thank the Heinz Endowment, all student interns, teachers and instructors of local school and summer camp, as well our collaborators Dr. Jessica Hammer, Dr. Louise-Phillipe Morency, Dr. Geoff Kaufman and Alexandra To for supporting the SCIPR project.

References


IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. XX, NO. XX, XX 2018

18


P. Jansen, N. Balasubramanian, M. Surdeanu, and P. Clark, “What’s in an explanation? characterizing knowledge and inference requirements for elementary science exams.”


Tanmay Sinha is a doctoral student in Learning Sciences at ETH Zurich. His research spans the investigation of individual learner behaviors in learning technologies, the study of social and interpersonal factors in learning, and the design of educational interventions to support failure-driven learning. He is interested in understanding learning mechanisms that bridge cognitive, socio-emotional and embodied levels, to develop learning technologies that can balance hard and soft skill acquisition.

Zhen Bai is a post-doctoral fellow in the School of Computer Science at Carnegie Mellon University. She is interested in designing interactive and intelligent interfaces that support positive development and lifelong learning for children with impaired social skills and disadvantaged backgrounds by augmenting social interaction and meaning making in a playful and collaborative manner.

Justine Cassell is the Associate Dean of the School of Computer Science at Carnegie Mellon University, and until recently was Director of the Human-Computer Interaction Institute in the School of Computer Science. Her research focuses on understanding natural forms of communication, and then creating technological tools for those forms of communication and linguistic expression to flourish in the digital world.