Learners' Reactions to Chatbot Communication Breakdowns: Insights into Fostering Learning

Ha Nguyen, University of California-Irvine, thicn@uci.edu

Abstract: Chatbots can promote productive discussion and learning. However, chatbots sometimes fail to interpret users' intent, leading to communication breakdowns and user disengagement. To alleviate these issues, researchers have experimented with designs that may increase users' willingness to fix breakdowns. This study explores the affordances of varying chatbot personalities (as a peer versus an expert) to prompt user repairs. Data sources drew from the chat logs of 18 small groups discussions in science. Results suggest that groups more often rephrased and provided explanations during communication failures with the peer chatbot than the expert version. These patterns, particularly providing explanations, had learning implications. Findings illustrate the affordances of designs that enact certain communication strategies—ones that may enrich learning during frictions in computer-supported interactions.

Introduction

Chatbots (text-based pedagogical agents) can facilitate discussions to promote productive talk moves and enrich learning (Dyke et al., 2013). These systems apply natural language processing of textual and speech input to provide scaffolds for idea articulation, knowledge construction, and argumentation (Kim & Baylor, 2016). However, researchers have documented cases where users ignore the chatbots' hints or abandon the conversations (Kumar et al., 2010). A potential reason for such disengagement is communication breakdowns, where the chatbots fail to interpret the users' intent and respond appropriately. To alleviate breakdowns, researchers have explored ways to improve the bots' natural language understanding and designed bot responses to increase users' willingness to correct the chatbots when the conversations fail (Adamson et al., 2014; Ashktorab et al., 2019; Engelhardt et al., 2017; Lee et al., 2010). Another strategy is to develop chatbot personas (i.e., personalities, characteristics) that may prime users' expectations for social interactions (Chaves & Gerosa, 2019; Kim & Baylor, 2016). In this study, I explored the strategies that learner groups employed to repair chatbot-facilitated discussions, and how these strategies differed in interactions with a less knowledgeable peer versus an expert chatbot. Understanding how groups approached breakdowns provides insights to enrich learning interactions.

Breakdowns, user repair strategies, and learning implications

Chatbots can serve as learning facilitators, to prompt learners who are less active to voice opinions, identify areas of misunderstanding, and ask for elaboration of peers' ideas to advance the group's knowledge (Dyke et al., 2013). Chatbots can also directly collaborate with learners, such as when learners critique bots' ideas or provide explanations to the bots to advance knowledge (Biswas et al., 2016; Graesser, 2016). It is important to examine group-chatbot collaboration, to avoid situations where groups abuse or ignore the chatbots (Kumar et al., 2010). I explore this question, with particular regards to how groups react to breakdowns in chatbot communication.

Breakdowns happen when the chatbots cannot comprehend user input and fail to complete the expected tasks (Feine et al., 2019; Følstad et al., 2018). Breakdowns can have detrimental impact on users' interactions, for example, users may reduce their trust in the chatbots or abandon the tasks altogether (Luger & Sellen, 2016). However, breakdowns also provide opportunities for learners to reflect on gaps in their understanding (Roscoe & Chi, 2007). When a chatbot misinterprets an utterance, learners can self-repair the misunderstanding or turn to others in the group to co-construct more coherent explanations. Researchers have documented a range of repair strategies in human-human and human-computer interactions (Ashktorab et al., 2019; Clark & Brennan, 1991). For example, users can rephrase the utterances or explain their intent (Clark & Brennan, 1991). Self-explaining or explaining to others can facilitate knowledge integration and learning (Chi, 2009; Roscoe & Chi, 2007).

Chatbot designs to facilitate user-initiated repairs

Chatbots can facilitate repair strategies through different talk moves. They can request that the user rephrase their utterances (e.g., "I don't understand. Can you rephrase?"), provide options (e.g., "Here are what I can help you with: Pick A, B, or C"), apologize (e.g., "Sorry, I don't understand"), and ask for feedback (Ashktorab et al., 2019; Engelhardt et al., 2017; Lee et al., 2010). Humanlike, social behaviors such as feedback requests would make the users more tolerant of breakdowns and more likely to initiate repairs (Engelhardt et al., 2017).

Researchers have experimented with varying personas (appearances, personalities, roles, etc.) to adapt to users' expectations for interactions with technologies like chatbots, voice interfaces, or robots. For instance, students gave and sought help differently in interactions with an expert versus a peer pedagogical agent (Biswas et al., 2016). Lee et al. (2010) found that users who perceived human-robot interactions as relational preferred apologetic repairs. Meanwhile, those who saw the interactions as transactional preferred compensation in the form of a coupon or a refund. It follows that minimal design cues may facilitate different user-initiated repair strategies. Building on research on different interactions with expert-peer chatbots in one-on-one interactions between a student and a bot (Biswas et al., 2016; Kim & Baylor, 2016), I explore the following questions in group settings:

RQ1. What reactions do learner groups have when the chatbot fails to address their intents sufficiently?

RQ2. To what extent do learner groups' reactions vary with different chatbot designs?

Methodology

The study took place within a high school science curriculum in southwestern United States. Students learned about marine protection regulations at a local state park, and collected and analyzed data at the park to understand how regulations might impact biodiversity. Early in the program, students interacted in small groups to build a concept map of the local marine ecosystem. The chatbots (nicknamed Kibot) were integrated into the activity to deepen discussion around ecosystems relationships. The chatbots kept track of the evolving concept maps to provide nudges in the same chat windows. The nudges focused on idea elaboration (e.g., "Can you explain more?"), systems connections (e.g., "If plastic pollution increases, what would happen to the habitat?"), and knowledge construction (e.g., "Do you agree with Erin? Why or why not?"). The two chatbots sent prompts with the same underlying functions, but differed in their appearances and speech styles.

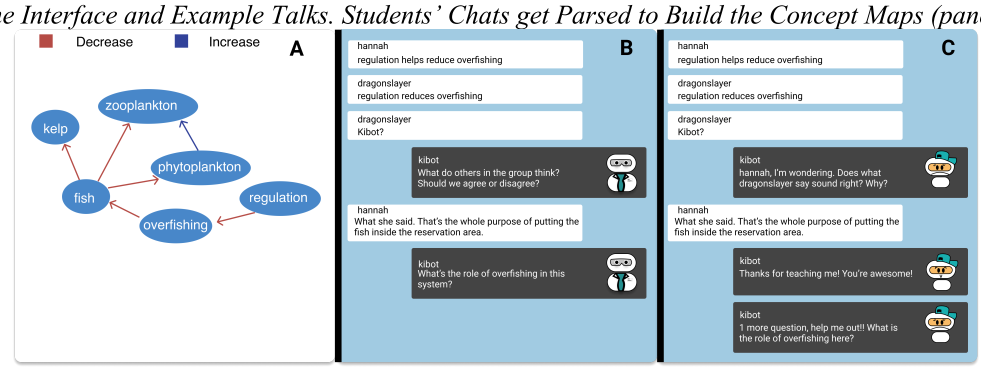
The less knowledgeable peer bot (panel C, Figure 1) resembled a peer with less knowledge about the marine ecosystem. This chatbot introduced itself as someone who had never been to the park and would benefit from learners' knowledge. The chatbot embraced dynamic expressions, such as frowning when confused. Meanwhile, the expert chatbot (panel B, Figure 1) introduced itself as someone with scientific expertise to test students' knowledge, and kept the same expression throughout. When learners asked the chatbots for hints, the bots would encourage group exchange (expert: "Ah, that's a simple question. Why don't we discuss with friends and answer it by yourself?"; peer: "Can you discuss and teach me?"). Figure 1 illustrates the bots' speech styles and an instance of breakdowns, where the bots could not parse the connection between regulation and overfishing.

Procedures

Participants involved 18 groups (52 learners) from two ninth-grade classes taught by the same science teacher in a public high school. The majority of participants were White or Hispanic. In a within-subject design, learner groups interacted with both chatbots (i.e., randomly start with one, and switch to the other design halfway through the lesson). Learners had no prior experience with learning chatbots. Because of the COVID-19 pandemic, learners were practicing social distancing in the classrooms and using the chat windows to communicate.

Figure 1

The Interface and Example Talks. Students' Chats get Parsed to Build the Concept Maps (panel A)



Data sources came from the group chat log, which was coded for repair strategies. The analyses were at the group level (involving 2-3 learners). To answer RQ1, I manually identified instances of communication breakdowns where a component or systems connection in the group chats did not appear in the system-generated concept maps. Table 1 presents the codebook, which was informed by prior work on user reactions to chatbots' failures (Feine et al., 2019; Kumar et al., 2010; Kvale et al., 2019). I achieved acceptable inter-rater reliability with a second coder on 20% of the data (average Cohen's $\kappa = .77$). To understand how repair strategies differed between chatbot conditions (RQ2), I ran Wilcoxon signed-rank tests (paired observations) to compare the repair strategies between chatbots. To account for multiple comparisons, I used the Benjamini-Hochberg procedure.

Table 1
Coding Scheme for Reaction to Chatbot's Failure to Interpret Intent

Codes	Definitions	Examples
Repeat (reframe)	Repeat a prior utterance	S1: Overfishing decreases fish. S2: Hello? Overfishing decreases fish.
Rephrase (reframe)	Reframe an utterance by changing the wording or adding words (without changing overall meaning)	S1: Overfishing decreases fish. S1: When humans hunt fish and fish too much, the fish population decreases.
Rephrase by imitating the chatbot's talk (reframe)	Reframe an utterance using similar sentence structures to the chatbots' nudges	Kibot: S2, what would happen if we overfish? S1: Overfishing decreases fish. S3: What happens to the fish if we overfish?
Rephrase and explain (reframe)	Provide additional explanations to clarify intent	S1: Overfishing decreases fish. S1: When we overfish, fish populations may not have as much time to reproduce.
Frustration (quit)	Express negative emotions	S1: Urgh, why doesn't it connect?
Change topics (quit, switch subject)	Switch to a new idea unrelated to the current ideas	S1: Overfishing decreases fish. S2: Phytoplankton increases zooplankton.

Results

Overall, 15% of the sessions included an attempt to fix the communication (271 instances out of 1,764 messages). For RQ1, I calculated the frequencies of repair strategies. The most frequent strategies were rephrasing ($N = 124$; 46%), followed by repeating ($N = 44$; 16%) and explaining ($N = 43$; 16%). There were few instances of imitating the chatbots ($N = 21$; 8%), switching topics ($N = 26$; 10%), and frustration ($N = 13$; 5%). Thirty-eight percent of the repairs happened when multiple individuals joined in to correct the chatbots. Consider the following excerpt:

S1: ocean acidification decreases biodiversity.
S1: does ocean acidification decrease biodiversity?
S1: how can I
Kibot: I'm learning so much from you! If there is something I did not catch, can you explain to me?
S2: ocean aci [acidification] decreases biodiversity because it kills a lot of species

In this excerpt, the students S1 and S2 were trying to build a connection between ocean acidification and biodiversity. Because this link did not exist in the underlying expert map, the less knowledgeable peer Kibot did not parse the connection correctly. S1 attempted to reframe the connection (posing a question), while S2 provided an explanation ("… because it kills a lot of species"). This excerpt illustrates how explaining behaviors can serve as a form of constructive knowledge building, where learners bring in additional reasoning (Chi, 2009).

The Wilcoxon signed-rank tests suggest differences in repair strategies between chatbot conditions (RQ2). Groups appeared to use more reframing and explaining in interaction with the peer chatbot, compared to the expert chatbot (reframing: M expert = 1.50, $SD = 1.54$; M peer = 4.09, $SD = 3.60$, $p = .01$; explaining: M expert = .80, $SD = 1.67$; M peer = 1.17; $SD = 1.06$, $p = .02$). This may indicate a higher tolerance for the less knowledgeable peer chatbot's mistakes and willingness to correct conversational breakdowns.

Discussion

Failure to respond to user intent and maintain dialogues is a common challenge that may hinder user engagement and trust (Kvale et al., 2019; Luger & Sellen, 2016). The low abandonment rate in the current work is thus encouraging, with evidence that learners most often rephrased, repeated, and elaborated on their ideas. Repeating information may not adequately promote learning by itself (Teasley, 1995). Prompting for users' explanations, utilizing rewards, and providing transparency in the underlying algorithm are promising strategies to facilitate users' repairs and responsiveness to the chatbots (Ashktorab et al., 2019; Lee et al., 2010; Nguyen, 2021).

Findings also illustrate that tweaks in chatbot appearances and speeches may increase users' willingness to engage with the technology (Ashktorab et al., 2019; Chaves & Gerosa, 2019; Yu et al., 2016). I found that student groups employed more reframing and explaining with the less knowledgeable peer chatbot, compared to the expert version. These findings can be linked to the learn-by-teaching paradigm, which creates opportunities for learners to seek information and identify gaps in their knowledge through testing an agent (Biswas et al., 2016). Explaining can offer opportunities to reason about concepts from alternative perspectives (Chi, 2009).

Overall, this study shows that unintentional moments of conversational breakdowns are not necessarily detrimental to learning. Results serve as a starting point to encourage future work on the interactions of designs, user preferences, and learning. For example, what should a chatbot look and act like when co-creating with student groups, beyond just providing nudges? What are potential frictions in group-chatbot interactions that can serve as opportunities for knowledge construction?

References

Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, *24*(1), 92–124. https://doi.org/10.1007/s40593-013-0012-6

Ashktorab, Z., Jain, M., Vera Liao, Q., & Weisz, J. D. (2019). Resilient chatbots: Repair strategy preferences for conversational breakdowns. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3290605.3300484

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain. *International Journal of Artificial Intelligence in Education*, *26*(1), 350–364. https://doi.org/10.1007/s40593-015-0057-9

Chaves, A. P., & Gerosa, M. A. (2019). How should my chatbot interact? A survey on human-chatbot interaction design. *ArXiv Preprint ArXiv:1904:02743*. https://doi.org/10.1080/10447318.2020.1841438

Chi, M. T. H. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/J.1756-8765.2008.01005.X

Clark, H. H., & Brennan, S. E. (1991). *Grounding in communication.*

Dyke, G., Adamson, D., Howley, I., & Rose, C. P. (2013). Enhancing scientific reasoning and discussion with conversational agents. *IEEE Transactions on Learning Technologies*, *6*(3), 240–247. https://doi.org/10.1109/TLT.2013.25

Engelhardt, S., Hansson, E., & Leite, I. (2017). Better Faulty than Sorry: Investigating Social Recovery Strategies to Minimize the Impact of Failure in Human-Robot Interaction. *WCIHAI@ IVA*, 19–27.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/J.IJHCS.2019.07.009

Følstad, A., Nordheim, C. B., & Bjørkli, C. A. (2018). What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11193 LNCS*, 194–208. https://doi.org/10.1007/978-3-030-01437-7_16

Graesser, A. C. (2016). Conversations with AutoTutor Help Students Learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124–132. https://doi.org/10.1007/s40593-015-0086-4

Kim, Y., & Baylor, A. L. (2016). Research-Based Design of Pedagogical Agent Roles: A Review, Progress, and Recommendations. *International Journal of Artificial Intelligence in Education*, *26*(1), 160–169. https://doi.org/10.1007/s40593-015-0055-y

Kumar, R., Ai, H., Beuth, J. L., & Rosé, C. P. (2010). Socially capable conversational tutors can be effective in collaborative learning situations. *Lecture Notes in Computer Science*, *6094 LNCS*(PART 1), 156–164. https://doi.org/10.1007/978-3-642-13388-6_20

Kvale, K., Sell, O. A., Hodnebrog, S., & Følstad, A. (2019). Improving Conversations: Lessons Learnt from Manual Analysis of Chatbot Dialogues. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11970 LNCS*, 187–200. https://doi.org/10.1007/978-3-030-39540-7_13

Lee, M. K., Kiesler, S., Forlizzi, J., Srinivasa, S., & Rybski, P. (2010). *Gracefully mitigating breakdowns in robotic services*. 203–210. https://doi.org/10.1109/HRI.2010.5453195

Luger, E., & Sellen, A. (2016). "Like having a really bad pa": The gulf between user expectation and experience of conversational agents. *Conference on Human Factors in Computing Systems - Proceedings*, 5286–5297. https://doi.org/10.1145/2858036.2858288

Nguyen, H. (2021). Exploring Group Discussion with Conversational Agents Using Epistemic Network Analysis. *Communications in Computer and Information Science*, *1522 CCIS*, 378–394. https://doi.org/10.1007/978-3-030-93859-8_25

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research*, *77*(4), 534–574. https://doi.org/10.3102/0034654307309920

Teasley, S. D. (1995). The Role of Talk in Children's Peer Collaborations. *Developmental Psychology*, *31*(2), 207–220. https://doi.org/10.1037/0012-1649.31.2.207

Yu, Z., Nicolich-Henkin, L., Black, A. W., & Rudnicky, A. (2016). A wizard-of-oz study on a non-task-oriented dialog systems that reacts to user engagement. *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 55–63.