Extending Chatbots to Probe Users:

Enhancing Complex Decision-Making Through Probing Conversations

Leon Reicherts
University College London
United Kingdom
l.reicherts.17@ucl.ac.uk

Gun Woo (Warren) Park
University of Toronto
Canada
warren@cs.toronto.edu

Yvonne Rogers
University College London
United Kingdom
y.rogers@ucl.ac.uk

ABSTRACT

Chatbots have become commonplace – they can provide customer support, take orders, collect feedback, and even provide (mental) health support. Despite this diversity, the opportunities of designing chatbots for more complex decision-making tasks remain largely underexplored. Bearing this in mind leads us to ask: How can chatbots be embedded into software tools used for complex decision-making and designed to scaffold and probe human cognition? The goal of our research was to explore possible uses of such "probing bots". The domain we examined was stock investment where many complex decisions need to be made. In our study, different types of investors interacted with a prototype, which we called "ProberBot", and subsequently took part in in-depth interviews. They generally found our ProberBot was effective at supporting their thinking but when this is desirable depends on the type of task and activity. We discuss these and other findings as well as design considerations for developing ProberBots for similar types of decision-making tasks.

CCS CONCEPTS

• Human-centered computing~Human computer interaction (HCI) ~Interaction paradigms~Natural language interfaces
• Human-centered computing~Human computer interaction (HCI) ~Interaction techniques

KEYWORDS

Chatbot, cognition, scaffolding, probing, decision-making, cognitive bias, conversational, proactive agent, agent intervention

ACM Reference format:

Leon Reicherts, Gun Woo (Warren) Park and Yvonne Rogers. 2022 Extending Chatbots to Probe Users: Enhancing Complex Decision-Making Through Probing Conversations. In *4th Conference on Conversational User Interfaces (CUI 2022), July 26–28, 2022, Glasgow, United Kingdom.* ACM, New York, NY, USA, 11 pages.
https://doi.org/10.1145/3543829.3543832

1 Introduction

Chatbots, being a popular form of Conversational User Interface (CUI), have become widely used for customer service applications (e.g. [9,49,50]). In particular, they have become commonplace on websites of businesses – helping customers find a suitable service or product, answering their questions, allowing them to make a booking or submit an order. While now more mainstream for such uses, their potential for other scenarios, such as to probe and guide users at the interface when completing more complex cognitive tasks has yet to be leveraged.

Here, we are interested in exploring how chatbots can be embedded into interfaces used for complex decision-making tasks. In other words, how can we design conversational interactions that integrate into ongoing thinking where users need to act out a series of decisions at the interface? Typically, interfaces for software applications are designed to help speed up and make task completion more efficient. Here, we are interested in designing the interface to encourage the user to "step back" and think about what they are doing – and to consider when this "slowing" down of interactions might be deemed advantageous. One potential benefit is to encourage the user to reflect more, especially when it is important to consider alternatives, rather making a rash decision. This raises the question, however, of how to achieve this while avoiding the chatbot adding too much "friction" causing the user to potentially get distracted and frustrated.

To address this question, we designed a form of chatbot interaction that was intended to probe and scaffold human decision-making in the moment it occurs. When the user is about to make a decision at the interface the chatbot is programmed to ask them particular questions that are meant to trigger critical thinking and help them externalise their thought process. Rather than chatting with the user, the chatbot in this decision-making context is meant to *probe* the user's thinking. Accordingly, we decided to name it a "ProberBot". To evaluate the ProberBot we chose an investment decision-making scenario. There are two main reasons for this. Firstly, it is a complex task that entails much uncertainty. Secondly, there is extensive research on (stock) investment decision-making that suggests that when investing in stocks people are prone to make quick and rash decisions they may regret later, which suggests this type of decision-making activity is a good candidate for being augmented by a ProberBot that is embedded into the interface where the decisions are made.

We designed an interactive decision-support bot prototype for this purpose, based on common cognitive and emotional biases in investment decision-making [4]. ProberBot was designed to be integrated in a trading software tool. The reason for this was to simulate real-world conditions that stock investors face when having to make a series of decisions based on what is happening in the market, for example, if stocks start falling unexpectedly. To evaluate the use of the ProberBot in situ we asked experienced stock investors to conduct several scenario-based investment decision-making tasks. Following this, we conducted in-depth interviews with them, asking them about their perceptions of and thoughts on using the trading interface plus ProberBot and, in particular, how it impacted on their thinking while using it and afterwards. They particularly liked the way it made them slow down for a moment during which they asked themselves questions concerning their decision in a structured way.

The primary contribution of the research reported here is to show how integrating a new kind of chatbot, called a ProberBot – because of the way it prompts and probes the user at certain times – in a trading tool/interface was found to be a valuable cognitive tool for investors; they thought it could help them better think through certain decisions which could otherwise be impulsive, driven by emotion, and potentially irrational.

2 Background

Beyond commercial uses, conversational user interfaces (CUIs) and chatbots have more recently begun to be explored for their potential benefits for other kinds of application areas. For example, some researchers have investigated their potential use for management decision-making [13]. Others have designed them to support collaborative idea generation [21] and to promote diverse discussions in group chats [22]. In addition, they are increasingly being used as an alternative to conventional surveys, since participants were found to not only write answers to open questions that are not only longer but also of higher quality [48]. They have also been used to conduct interviews due to their potential to empathically handle users' open-ended questions and non-linear dialogues [17,47]. One main focus in Human-Computer Interaction (HCI) in recent years has been on designing them as interventions for (mental) health and wellbeing [3,23–25,27,29,31,32,34,40–42]. Some of the emerging findings suggest that they have a number of benefits when applied to this domain, including making it easier for people to disclose sensitive information to get support, in particular, for issues that may involve social stigma or supporting wellbeing through promoting self-reflection and self-compassion techniques.

CUIs Supporting Learning. Another popular area for which CUIs have been designed is as interventions to improve education and learning [12,15,16,26,37–39,46]. In particular, they have been designed to guide learners through learning tasks, scaffold their thinking, and test them on their knowledge. For example, Alaimi et al. [1] developed a CUI that fosters question-asking skills in children and Wambsganss et al. [43] developed a conversational agent that provides feedback on students' argumentation when doing a persuasive writing exercise. They have also been designed to foster metacognitive skills such as think-aloud, which resulted in improved learning outcomes [35]. An interface for online video lectures, was proposed by Winkler et al. [45]. The CUI/agent asked questions at specific points about the content learned in the lecture. If a learner's answer to the agent's question was incorrect, it would follow-up with sub-dialogues that stated the problem and question in different ways and by that scaffolded the learner's thinking and guided them to the correct answer. Another example within an educational context but with a different goal is *Muse* by Cabales [7], which prompted students to monitor and reflect on their learning strategies while working on research projects. In a small user study, the agent's scaffolding of metacognitive reflection appeared to have helped students apply beneficial learning strategies, indicating possible metacognitive behavior change. This body of research shows the diverse uses of CUIs for learning, where CUIs have been successfully used to guide learners, support question-asking and argumentation skills and facilitate metacognitive processes, for example to help them adhere to more beneficial (learning) strategies.

CUIs Supporting Financial Decisions. In the financial services, chatbots have been adopted for common tasks like providing support to customers or allowing them to execute certain orders and transactions. One popular use case in investing contexts has been to use conversational "robo-advisors" [8,11,19,28]. Such advisors often aim to (i) capture user's interests and preferences (e.g., for specific industries), their values (for example concerning environmental, social, and corporate governance aspects), and their risk aversion, to then (ii) provide suggestions for suitable financial products, investment strategies, as well as portfolio allocations. For a financial service provider, an advantage of using such robo-advisors is to increase the efficiency of their customer onboarding and customer support processes. A key benefit for the customer can be a more seamless onboarding experience as well as a swift and user-friendly provision of individualised investment suggestions. Here we are interested in a chatbot that supports and interacts with the user/investor beyond the guidance of current robo-advisors concerning which products or strategy to choose. Instead, it aims to probe and scaffold the investor's decision-making at specific moments to help them *stay on course* with their strategy and avoid emotional (re)actions. Our motivation for doing this is to draw from the research in other domains, including learning, where conversational interactions were found to not only be capable of scaffolding the user's cognition but also have potential for engaging them in self-reflection and, in particular, metacognition (i.e., awareness of one's own thought processes) which can enable more strategic thinking and improve decision-making.

3 Method and Research Questions

We adopted a method inspired by the *technology probe* approach developed by Hilary Hutchinson and colleagues 20 years ago [20] that has since become a mainstream prototyping technique in HCI. Essentially, a partially functioning prototype is built with the aim

of collecting data about its use in/through a real-world scenario or context. Here, we use the method to discover more about how probing conversational agent interventions will be perceived and experienced by users, in particular, how they think these kinds of probes will affect their decision-making and reflections surrounding this. Using the technology probe prototype, we ask the following related research questions:

- *The Tool:* How should a probing conversational interface (called a ProberBot) be designed to be embedded into a software tool that will encourage metacognition during decision-making?
- *The Domain:* How do expert investors – familiar with online investing platforms – think about having a ProberBot added/ integrated into a trading tool, which asks them questions regarding their investment decisions? Does it help or distract them when considering what to do with each investment?

4 ProberBot Design

The rationale behind the concept of ProberBot is that it can proactively intervene when a user is about to make an important and potentially risky decision while interacting with a software tool (in this case a trading tool). The design of ProberBot was informed by well-known cognitive and emotional biases that occur in investment decision-making (e.g., [4,51]). These biases have been found to often lead to poor decision-making in investing, negatively affecting the investing performance/returns [6]. We started the design process by creating ProberBot dialogues which were based on some of these biases and which intended to address them; we then evaluated these dialogues with experts to see how they could be improved and to decide which of them to use in the subsequent study. We describe the steps of the design process in more detail in the following sections.

4.1 Dialogue Design for the ProberBot

Three common biases of investment decision-making were considered as a basis for the ProberBot dialogues. The biases we chose to inform the design of our ProberBot dialogues were (i) the *hindsight bias* and (ii) the *availability bias and* (iii) the *disposition effect* (which also represents a *bias* even if it does not carry the word "bias" in its name). The *hindsight bias,* also known as the "knew-it-all-along" effect [14,36], refers to overestimating the ability of oneself to have predicted an outcome after it has already happened. As a result, the hindsight bias can lead (together with other factors) to people being overconfident, since they believe that they have accurately predicted past events. *Availability or recency bias* refers to overestimating the importance of recent events (e.g., news) and underestimating other information [2,10]. The *disposition effect* refers to a common tendency of investors of holding "losing stocks" ("losing" in the sense that the stocks have been losing value since they were originally purchased, meaning that the current price is lower than the initial purchase price) too long while selling "winning stocks" (meaning that the current price is higher than the initial purchase price) too soon [44].

The idea was to design ProberBot dialogues that could potentially improve upon this situation by being triggered in situations where there may be a bias in the user's decision-making, helping the user realise that their decisions could be biased (such as keeping a losing stock without a clear reason to do so).

Until now, there have been various strategies proposed to help investors reduce the effect of biases creeping in their thinking. These include following a clearly defined analytical process that can be tested and retested and adjusted throughout time [5,30,51] while keeping track of personal decisions and mistakes and successes by keeping a "investing journal/diary" and other accountability mechanisms (see also [18]). However, all of these are time-consuming and difficult to maintain. They may help with record keeping and reflecting but they are not as suitable for/effective at supporting in-the-moment decision-making. One of the benefits of our proposed chatbot that is embedded into the trading platform interface is that it can potentially reduce the risk of certain biases head-on by encouraging the user to engage in meta-level thinking when it matters most, notably during the ongoing decision-making process. Part of what we are interested in is whether this can trigger an "inner dialogue" that makes parts of the decision-making clearer. This process of externalisation should make vague thoughts more explicit and bring to light inconsistencies in argumentations and decisions.

4.1.1 Designing the Initial Dialogues. The dialogues were designed to *indirectly* address the biases, rather than trying to avert them specifically. The reason for this approach was that although *general* patterns in people's behaviour can indicate the presence of a bias (such as systematically holding losing stocks longer than winning stocks) it is often difficult to determine if a bias is indeed present in a *specific* decision. The intention is thus to encourage the investor to think more about their decision in the moment and help them become aware themselves when they might succumb to one of the biases. The dialogues we designed related to the three biases were the following:

Dialogue 1. *Part A: Asking the user to formulate an investment thesis/motivation when making a major buy decision. Part B: Reminding the user of their initial/previous investment thesis and asking them to what extent it still holds if they intend to sell the same stock soon after initial purchase (i.e., Part A).*

This dialogue asks the user to indicate the extent to which they think their previously defined investment thesis for a stock still holds when they intend to sell it soon after having purchased it. It also asks for the factors which made them change their mind from the initial investment thesis.

Dialogue 2. *Asking the user for the reasons for still holding a stock which has been falling for an extended period of time.*

The dialogue starts off by asking the user how likely they think a recovery of the stock price is (i.e., to the price level at which they initially purchased it) and then asks them to provide the main reasons for their estimation/evaluation. It then gets the user to formalise the potential risks and gains by formulating a best- and worst-case scenario and rating the probabilities.

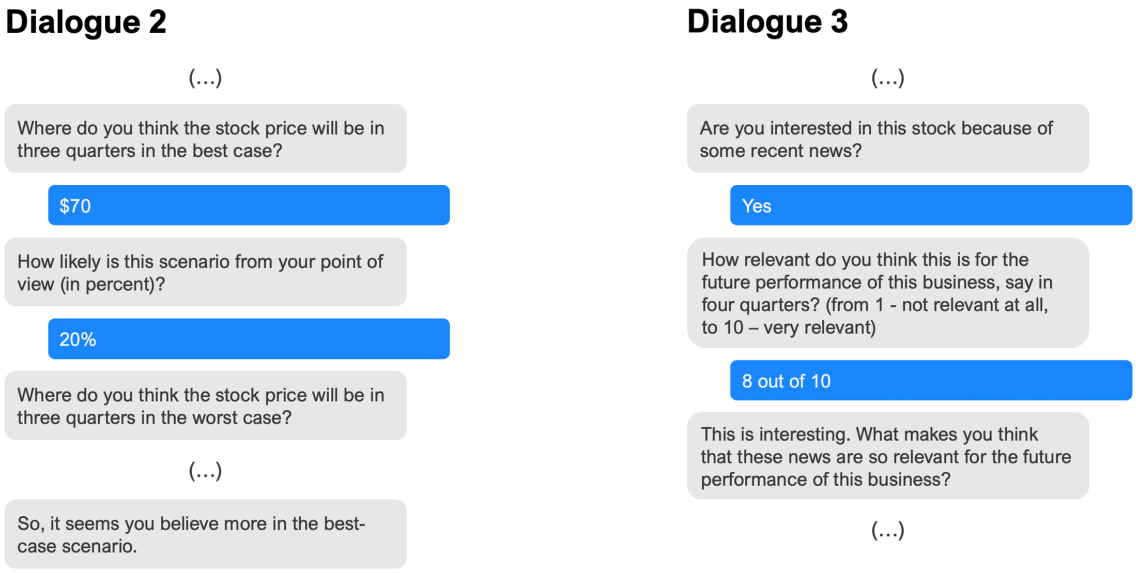

Figure 1: Excerpts of Dialogue 2 and Dialogue 3 which probe user's thinking with context-dependent prompts/questions.

Through this, it is intended to help the user become aware of the potential opportunity cost (i.e., the cost of not selling that losing stock/replacing it with another). In the end it presents back a summary of the user's evaluation.

Dialogue 3. *Asking the user how relevant they consider a recent news item for the future of a stock.*

On the one hand, company-related news can be selective and exaggerated and on the other they are usually quickly "picked up" by the market and reflected in the stock price (this can often happen faster than an individual investor will notice the news and be able to respond to them). To reduce the risk of a potentially impulsive decision, based on attention-grabbing news, this dialogue asks users to reflect on a news item and its relevance.

Dialogue 1 (part A and B) is based on/informed by the hindsight bias; *Dialogue 2* is based on the disposition effect, and *Dialogue 3* is based on the recency/availability effect. As can be seen in Figure 1, the questions ProberBot asks do not give any specific directives but are intended to get the user to think about certain aspects, such as how likely they consider certain scenarios (an excerpt of Dialogue 1 can be found in Section 4.3).

The dialogues were designed to promote short conversation turns to provide a scaffold for getting the user to make specific estimations and evaluations (such as how likely they consider a certain future event). This was inspired by other chatbots that provide certain choices for people to choose from when self-reflecting, such as the mental-health app *Woebot* [33], where simple options/inputs are often sufficient to trigger the "inner dialogues" involved in self-reflection.

4.2 Expert Feedback

To explore further the interface's suitability of addressing these three biases through having a dialogue at the interface we approached two experts in the field, both having extensive professional stock investing experience. The dialogue "prototypes" for the three biases were presented to the two experts (using a digital whiteboard) for their feedback. A key comment they both made was that the ProberBot was too "pushy". It was argued that the ProberBot should just scaffold investors' thinking rather than be set up to ask "leading" questions or to make suggestions (due to the challenge of determining if a bias is in fact present as they also pointed out). For example, they critiqued a part of the disposition effect dialogue, which initially was "*That's interesting.* ***Maybe you want to reconsider*** *why you keep holding this stock?"* and suggested changing it to *"That's interesting.* ***So, you are currently holding a stock which you don't believe will recover.****"* This nuanced shift in expression makes the ProberBot seem less pushy, appearing as a more neutral statement, leaving it up to the user what they conclude from it and how they want to act upon it. Based on their feedback, similar changes were made to the phrasing of the other dialogues to make them more neutral as well, so that they would only provide "cognitive scaffolds" instead of hinting at or suggesting what the user should do.

One aspect that was particularly appreciated by both experts was that the data captured in ProberBot's dialogues could provide a useful history of decisions that users could refer back to, akin to keeping an investing diary (just in conversational form), allowing them to see and revisit what they decided previously and why.

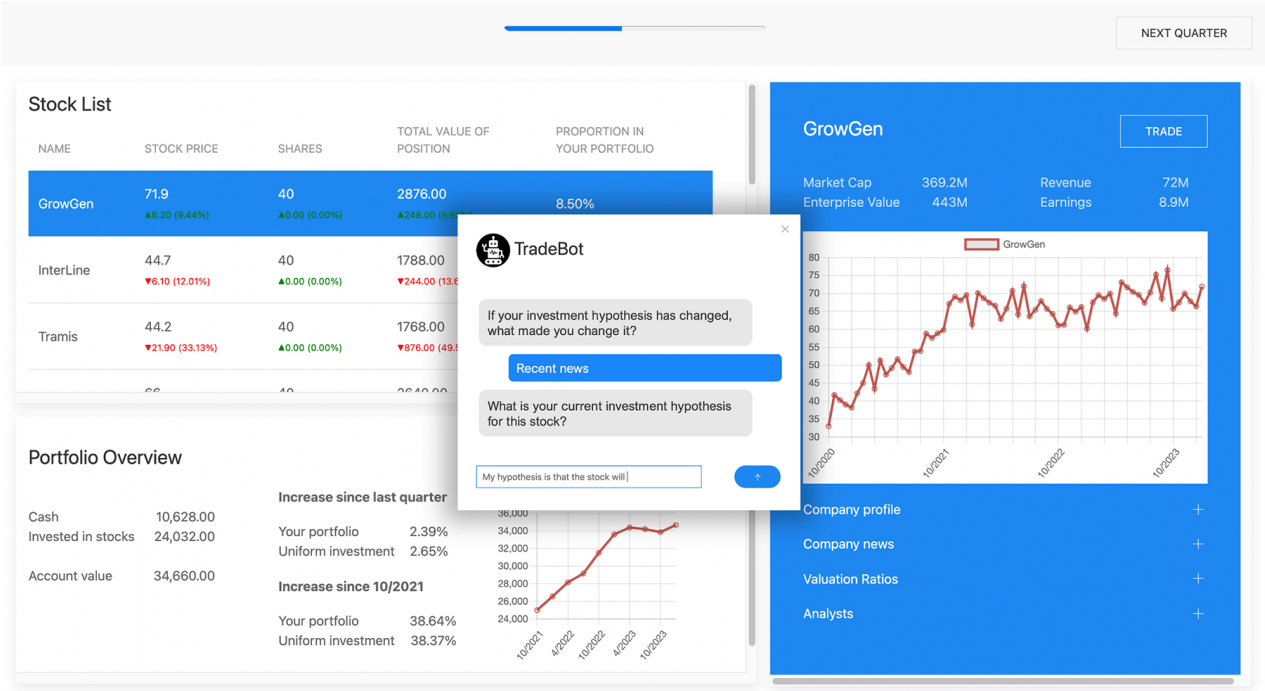

Figure 2: Trading simulator interface. On the left at the top is a list of all the stocks, underneath is an overview of a person's overall portfolio (including the performance throughout time) and on the right is a graph that shows the stock price over time. The information on the right would update when a stock is selected from the list. The ProberBot appears in a pop-up window in the middle at key moments when the user is about to make a trade.

Another general comment they made was that it can be very useful to have "someone" (i.e., this bot "entity") to talk to when thinking about their decisions, since it would get them to externalise their thoughts. Based on the experts' feedback, the dialogues were revised so that they would mainly get people to externalise their thought process and in so doing enable them to identify potential risks and biases by themselves rather than having the bot too explicitly referring to or hinting at certain biases.

4.3 Design of the Technology Probe

4.3.1 Trading Tool. We built a bespoke stock trading tool/simulation in which the ProberBot was embedded (see Figure 2 for a screenshot). Both the trading tool and ProberBot were designed to be authentic looking by providing a realistic level of functionality and interactivity. Various types of data that are common in trading interfaces were provided including stock price time series graphs, price/earnings ratio, price/book ratio, price/sales ratio, analyst ratings, consensus/average target price, market cap, revenue, company information/profile, and company news items. The trading tool was designed to provide five years of synthetic stock data, with a "Next Quarter" button that could be clicked on. Clicking the button would update the stock prices and market situation/context. Due to our focus on longer-term investing strategies (financial) quarters were chosen as a common and meaningful time interval in which an stock investor with a long-term strategy may consider and assess new investments, as well as their existing investments and portfolio performance.

The interface shows the stock price data and relevant metrics for three companies which were intended for our study scenarios. We also designed the combined trading tool and ProberBot so that rules could be set for each dialogue for when it should appear (e.g., when there would be volatility or certain trends in the stock price, duration of holding the stock etc.). For the present study the dialogues were "hard-coded" to appear at specific moments (as part of our scenarios) at which the user would intend to trade a specific stock. This was to assure increased control and comparability of the collected data. After interacting with the ProberBot the user has the option to move on to confirm or cancel their intended trade.

4.3.2 ProberBot Interface. The chatbot was designed to support three kinds of responses (see Figures 3-5): (i) text inputs, (ii) discrete scale inputs (1-10 or 0-100%), and (iii) multiple-choice inputs depending on the question asked. The choice of the response/input type depended on the question being asked and usually there was a mix of them in each dialogue. This enabled variability for the participants when considering how to react and reflect upon them. The chatbot dialogues were designed to have several branches using a determination logic based on predefined, context-dependent criteria. This allowed the chatbot to output context-aware responses based on previous user inputs, which were mostly follow-up questions or summaries of previous inputs.

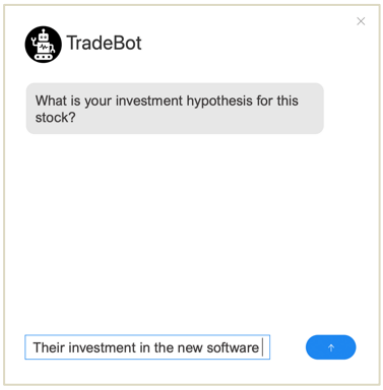

Figure 3: Text input example (from Dialogue 1 – part 1).

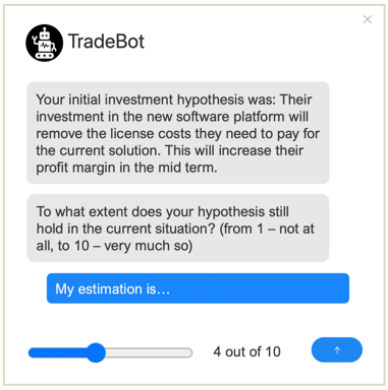

Figure 4: Scale input example (from Dialogue 1 – part 2).

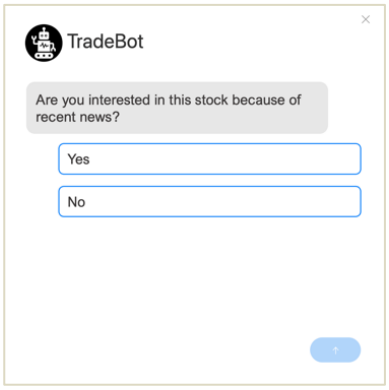

Figure 5: Multiple choice/yes-no example (from Dialogue 3).

5 Study Design

To evaluate the effectiveness of embedding various ProberBot dialogues in the trading software tool in terms of how they could support decision-making we designed an interactive scenario-based study. The reason for this design where participants interact with the tool themselves was that they would experience the tool and get a sense of how such an embedded bot "works". Our focus in this study was not on the ProberBot's usability or performance or on how exactly it changes or affects people's decisions. Instead, our focus was on (a) whether experienced investors would think it could help mitigate certain biases based on specific, realistic situations and decisions, (b) understanding the types of thought processes it can trigger (particularly metacognitive ones) and (c) how disruptive or intrusive the different ProberBot dialogues may be perceived. Furthermore, with the chosen study design we intended to get insights into people's understanding and conceptualisation of ProberBot and elicit their ideas on other cognitive tasks they would like a (future) ProberBot to help them with (and which *not*) based on their experience of interacting with it. To provide participants with a realistic experience of a possible use of ProberBot we designed a series of scenarios for which they had to work out what to do under the guidance of the researcher. The scenarios were: (i) investing in a stock which would subsequently have to be revisited/re-evaluated after performing badly over several quarters, (ii) evaluating their portfolio after a certain amount of time and deciding which stocks to continue to hold, and (iii) deciding about an investment in a stock which is currently being hyped by considering the different news items and other information and metrics regarding that stock. For all these scenarios participants interacted with the bot and the respective dialogues it provided. The researcher provided participants with certain hints and suggestions for their actions at specific points, since (a) we could not expect them to go through all the data and make a decision in the available time frame; (b) we wanted each participant to experience the same ProberBot dialogues to be able to better compare their reactions and thoughts. Ethics approval was obtained from our university prior to the study.

5.1 Participants

Six participants who had various levels of experience in investing in the stock market were recruited for the study. Four participants had professional trading experience in different contexts (e.g., working for stock markets or investment banks) and two participants had long-term private/retail investment experience. For this study, we intentionally only recruited experienced stock investors, since understanding the trading tool (including the different types of data it provided) and the ProberBot dialogues (and why they appeared) required sufficient investing experience.

5.2 Procedure

Participants were informed about the purpose of the study and asked to fill in a consent form agreeing to being audio and video recorded during the study for subsequent analysis. After a walkthrough of the trading tool and a familiarisation phase the researcher provided the first scenario. The interaction with the trading interface was guided by the researcher but was operated by the participant on their computer/browser while they were sharing their browser window through the video conferencing software with the researcher.

The scenarios provided in the study involve significant changes in companies' financial results (e.g., their earnings) and outlooks and their stock price after several financial quarters to imply situations of increased uncertainty in which the ProberBot could be triggered. After each interaction with the ProberBot, we asked the participants to guess why it was triggered, as well as whether they thought the questions/interactions were appropriate for the given situation. After their interactions with the ProberBot for each scenario were completed, the participants were interviewed about what they thought of having a probing chatbot embedded into a trading interface, their views on the dialogues it provided, and the extent of its intrusiveness and potential disruptiveness.

5.3 Data Analysis

The data collected was transcribed by one of the researchers. This covered statements made by participants while interacting with the tool and ProberBot and thinking aloud as well as the in-depth interviews following the investing task. The transcriptions were reviewed by a second researcher using an iterative and open form of thematic analysis. The main researcher went through all the interviews and inductively coded them. The other went through all the interviews and suggested changes to the codebook and the coding. Subsequently, all disagreements were discussed and resolved between the researchers. The themes were organised into (i) perceptions about the value of a ProberBot, (ii) challenges of having a ProberBot, (iii) expected use of a ProberBot and individual needs.

6 Findings

Taken together, all six participants had a similar understanding of what ProberBot is trying to achieve; they generally thought that its probing interactions could enhance their own decision-making by preventing certain impulsive actions or inconsistencies in decisions. However, some of the participants also had concerns about whether they would find certain dialogues beneficial when they are making investment decisions. These and other findings are presented in more detail in the following sections.

6.1 Perceptions About the Value of a ProberBot

When the participants were asked what they thought the purpose was of ProberBot, all of them mentioned its potential value of trying to support their decision-making when there is risk of (re)acting inadequately (e.g., not in line with previous investing decisions and strategy). Participants could see for most questions "where the ProberBot is coming from" or why it asks these questions. For example, Participant 6 mentioned:

> *It made you re-evaluate your thought rather than having a knee jerk reaction or emotional response (…) so it gives you a little bit time to think. (P6)*

Most participants also mentioned that they could see how it could help investors keep track of their own past decisions to enable them revisit or – if needed – revise their investment theses and/or strategy (which was a point particularly related to Dialogue 1):

> *It's nice to be reminded of your earlier decision-making process, because sometimes you forget you've actually done quite a lot of due diligence and thought about something very carefully. (P1)*

When asked about what they think could be triggers for ProberBot to appear, their explanations were very similar. For example:

> *You could interpret the situation for these companies as being they were particularly, uhm, not risky decisions necessarily, but where I needed to stop and think about it. It wasn't just a day-to-day kind of trade. (P5)*

6.2 Challenges of Having a ProberBot

Although all participants understood the value of having ProberBot intervene while they were trading, and were generally in favour of it, they also made comments about certain challenges. For example, two participants were concerned it could interrupt them and be intrusive:

> *It's a little bit intrusive but it's something I think you could get used to if you could make it almost part of the [personal investment decision-making] process. (P1)*

Another participant found the ProberBot to be "quite curt". Two participants mentioned that some dialogues may not be that helpful for them given their experience and their investing approach. A more general problem was pointed out by one participant regarding Dialogue 1, which was about capturing the initial investment thesis and then re-evaluating it. They mentioned that people would usually find a way to explain their previous actions afterwards ("post hoc") and then proceed with their decision even if it contradicts their initial investment thesis or their general strategy:

> *You know my challenge with a bot like this is it's forcing me to provide post hoc explanations, right. I've already decided what I'm going to do, now it's asking me to justify it. (P2)*

Dialogue 3, which asked the user about the relevance of a news item for a specific stock/company, was perceived to be less useful by three participants as they said they would usually not trade based on the news in general. One participant had even stronger feelings about it, but for a different reason, for them Dialogue 3 was going beyond just scaffolding their decision, leading them into a certain way of thinking:

> *Is the bot prompting me in a particular way to influence that train of thought somehow? (…) This feels more like it's no longer being a neutral bot, but it's actually leading me down a particular [direction]. So, previously [i.e., in previous dialogues] it was a reflection but this one really felt like it's pushing me in this direction of making particular assumptions or decisions. (P5)*

Participants also made several comments about the additional time that the interaction with the bot added to the decision-making. However, this comment was made largely when they were considering traders with short-term strategies (e.g., day trading), acknowledging that this would not apply to long-term investing strategies. One participant describes a dilemma of sometimes "not wanting but at the same time needing" the bot when in a rush and thus being at risk to act against one's long-term strategy or goals:

> *So now you need to click through it and some time passes, and if there's something which is really time sensitive so, for example, if I want to trade it now and I don't want to waste time telling the bot, why I wanted to buy or sell now, then I think I should have the ability to skip it. So maybe this is when the market is moving really quickly, but at the same time, maybe someone else might say, this is exactly when you need it, so you don't overreact. (P4)*

In other words, what P4 points out here is that some bot interactions can feel disruptive and slow the user down but that this may be exactly what is appropriate in certain situations where an investor may get emotional, impatient, or rushed.

6.3 Expected Use and Individual Needs

Participants had several ideas for further situations when they would use ProberBot and how it could prompt them with certain questions including what could trigger the questions to appear (e.g., certain user traits, states, or behaviours). For example:

> *I would ask the bot to ask me whether I'm really sure I want to buy or sell. After 8:00 o'clock at night I might have had too much to drink (...) I think that when the market dropped with COVID it would have been very useful to have a bot saying. "Are you sure [you want to sell]? Oh, go away and think about that." (P6)*

Although this was a somewhat humorous suggestion, the underlying idea of the bot probing them about their emotional and cognitive state and current context was also made by two other participants.

Participants also mentioned that ProberBot could help them better keep track of and make sense of their own performance and decision-making processes, for example:

> *Was the reason [for a previous "inappropriate" decision] I did not do my due diligence properly or did I do something different this time or was it 2:00 in the morning. If you could reflect that back that would be very useful. (P1)*

Four participants mentioned that a ProberBot should have certain options to control it. For example, P3 said:

> *Sometimes you go against your own nature [knowingly] (...) and a bot kindly reminding me, then, maybe I don't care what the bot says. Like, I know what he's going to say because I've used him for months now, and I know his nature so it's almost like you know, sometimes you just don't want to hear it. I'm going to trade this regardless (...) You could have like three levels of intrusiveness right like "standard", "super helpful", or "on the sidelines".*

Another observation concerned the gender of the ProberBot. P3 and P5 referred to the bot with "he/him" pronouns although we introduced it without a specific gender, the others referred to it in a gender-neutral way.

Finally, it was noted that investing decisions often depend on a variety of factors in addition to those currently reflected and considered in the bot dialogues, such as desired portfolio proportions/allocations (e.g., due to a specific strategy) that should be achieved or maintained. For example:

> *Was this – even if it's just trading one stock – a response to a market condition or is it in response to analysis of the portfolio, overall, or is it based on this instrument in you know, in a silo. (P3)*

The fact that (investment) decisions are often interdependent, as pointed out in this quote, is a key consideration for the design of ProberBots. As such, the dialogue design needs to take into account that decisions can build upon each other (i.e., depend on past decisions and affect future decisions) or they can be sub-decisions of a higher-level decision (e.g., the decision to sell a specific stock due to the higher-level decision to rebalance the portfolio in a specific way, as P3 alluded to in the above quote). Translating these interdependencies into equally (or at least comparably) interdependent dialogues which are then triggered at appropriate decision-making points is a key design challenge of building ProberBots – among various other challenges and considerations which we will outline in the next section.

7. Discussion

Taken together, the findings suggest that having a ProberBot embedded into trading tools/platforms can help a stock investor's decision-making by scaffolding aspects of their cognition. By this we mean the ProberBot has the effect at certain points to slow down the stock investor to stop momentarily and reflect on whether their next intention is wise. This seems most preferable for trading tasks that are not overly time-sensitive or following short-term strategies (as for example in day trading), where rapid decision-making is critical. For long-term investing strategies it was appreciated for being able to question their motivations, helping them to think more critically about their decisions – which they may not do when acting "in the moment". The participants did not raise specific concerns about the time the interaction with the bot may add to their decision-making process, but rather that this added thinking time is often even desirable for investors with long-term strategies.

However, the study also revealed certain challenges of using a ProberBot in this manner. For example, some participants thought that a ProberBot could be intrusive or that its dialogues may sometimes seem not relevant or effective. This implies that one of the difficulties of integrating a ProberBot into existing software tools, as we envisioned, is the risk of disrupting ongoing decision-making. Users may require some time to get used to this type of probing dialogue to overcome the perception it is getting in their way. Hence, it is not straightforward as to how to design the ProberBot dialogues so that they appear/occur at opportune times. They need to be sensitive to different user needs and when best to intervene. The user could be involved in helping here, by suggesting where and when they would like their ProberBot to intervene. This could include "telling" the ProberBot, in the set-up phase, their personal level of experience and investing strategy and then, when in use, having specific settings for how much the bot should intervene (e.g., "standard", "super helpful", or "on the sidelines" as suggested by one participant).

Having such controls could not only make the ProberBot more tailored to the users' needs but also make users feel comfortable to receive its prompts/probing dialogues from an early stage of usage, as they would be able to anticipate to some extent when and how certain dialogues would be triggered. This could potentially also result in a higher willingness to interact and engage with the ProberBot and consider its probing questions in their decision-making.

There were a few instances where participants were not sure of the benefits of specific ProberBot dialogues. For example, three participants mentioned Dialogue 3 appeared not relevant to how they usually make their investment decisions. The same dialogue was also perceived by one participant to be pushing them into a certain direction rather than supporting and scaffolding their decision-making. These comments suggest that the phrasing of ProberBot dialogues needs to be carefully crafted so that it is seen as being neutral and not perceived to be nudging the user to make a particular decision that they may not agree with. Furthermore, our findings also suggest that it may not be equally straightforward to define dialogues that users find helpful for all biases and that only certain ones are suitable to be addressed by a ProberBot.

The question was raised as to how the ProberBot could be designed, given it would not always have the relevant information for why the user is making a certain decision (e.g., someone selling certain stocks due to having to urgently pay back a student loan and not because they do not believe in the stock anymore). However, participants pointed out it may be difficult to know how best to configure when the ProberBot should *not* intervene. A possible solution is that the user would have to "explain" to the ProberBot why they would like to turn it off for certain transactions or for a certain period of time before being able to do so. However, allowing the user too much control in setting the dialogues could undermine the main purpose and the effectiveness of them as neutral "probes" that are intended to make stock investors reflect on their decisions in the moment and to reduce the risk of being impulsive and regretting it later.

In sum, our findings suggest that ProberBots are a promising approach for extending the remit of chatbots into helping human decision-making at or within the respective interface used for decision-making. An important finding from our research, however, is that it is not straightforward designing the kinds of dialogues for this. Even though we tried to design our dialogues to be neutral, not suggesting or implying a specific action/decision, they were sometimes perceived by some participants to be "leading". Other concerns include how to design them so as not to be too intrusive or distracting. There is a danger that they become seen as the new Microsoft *Clippy* and people turn them off before working out how they can be configured to help their decision-making in ways that benefits them.

Due to the focus of this study on getting insights into the types of (metacognitive) thought processes that a ProberBot can trigger and stock investors thoughts on, conceptualisation of and needs for such a bot, we have chosen a purely qualitative approach. Furthermore, the ProberBot and user actions occurred within the bounds of a predefined scenarios provided by the researcher and the synthetic data in the trading interface; although common scenarios were chosen that have a high degree of ecological validity it is likely that users' real-world usage of, response to, and experience with the ProberBot is different. Future work should thus investigate further how and to what extent certain dialogues and prompts can affect people's decision-making and the effects this may have on task performance (i.e., portfolio performance metrics) – ideally over time. Moreover, further research is needed to investigate how different kinds of dialogues can be used in the real world and tweaked to be at a level that stock investors find helpful. Our future research will also investigate ProberBots' effectiveness when in use by investors "in the wild" as well as further explore other domains and types of tasks where human decision-making would benefit of such conversational probing and scaffolding.

8. Conclusion

Our research has revealed the potential benefits of designing a chatbot embedded into a software tool used for decision-making to support decision-makers. This approach is quite different from the design of many chatbots which are intended to provide services and information to users through answering users' questions. We explored how it is possible to augment human cognition by considering a ProberBot's suggestions and thinking about what they are saying at a given time, and whether the user stays with the original intention or changes their course of action. Hence the approach we are advocating here is how designing a chatbot to be more of a *probing* bot can enable the user to think about their decision-making in the moment – especially in situations when there is a risk of impulsive and potentially biased decisions. However, to design probing dialogues and to determine when they appear is difficult, because stock market investors are often involved in demanding cognitive work when making decisions, meaning it could only require little to be disrupted. Enabling users to be involved in setting up the type and timing of the interventions could help them in finding a sweet spot for ProberBot engagement that suits them. Overall, ProberBots seem to be a valuable cognitive tool for (investment) decision-making.

ACKNOWLEDGMENTS

We would like to thank everyone who participated in our study and everyone who gave us valuable input and feedback for building the ProberBot prototype. We would like to extend our sincere thanks to the anonymous reviewers for their valuable suggestions on improving the manuscript. This work was partially funded by the Leverhulme Trust award DS-2017-026.

REFERENCES

[1] Mehdi Alaimi, Edith Law, Kevin Daniel Pantasdo, Pierre-Yves Oudeyer, and Hélène Sauzeon. 2020. Pedagogical Agents for Fostering Question-Asking Skills in Children. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), Association for Computing Machinery, Honolulu, HI, USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376776

[2] Vicky Arnold, Philip A. Collier, Stewart A. Leech, and Steve G. Sutton. 2000. The effect of experience and complexity on order and recency bias in decision making by professional accountants. *Accounting & Finance* 40, 2 (2000), 109–134. DOI:https://doi.org/10.1111/1467-629X.00039

[3] Petter Bae Bae Brandtzæg, Marita Skjuve, Kim Kristoffer Kristoffer Dysthe, and Asbjørn Følstad. 2021. When the Social Becomes Non-Human: Young People's Perception of Social Support in Chatbots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–13. DOI:https://doi.org/10.1145/3411764.3445318

[4] H. Kent Baker and John R. Nofsinger. 2002. Psychological Biases of Investors. *Financial Services Review* 11, 2 (Summer 2002), 97–116.

[5] H Kent Baker and Victor Ricciardi. 2014. How biases affect investor behaviour. *The European Financial Review* (2014), 7–10.

[6] Brad M. Barber and Terrance Odean. 2013. The Behavior of Individual Investors. In *Handbook of the Economics of Finance*. Elsevier, 1533–1570. DOI:https://doi.org/10.1016/B978-0-44-459406-8.00022-6

[7] Victoria Cabales. 2019. Muse: Scaffolding Metacognitive Reflection in Design-Based Research. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI EA '19), Association for Computing Machinery, New York, NY, USA, 1–6. DOI:https://doi.org/10.1145/3290607.3308450

[8] Heloisa Candello, Claudio Pinhanez, David Millen, and Bruna Daniele Andrade. 2017. Shaping the Experience of a Cognitive Investment Adviser. In *Design, User Experience, and Usability: Understanding Users and Contexts* (Lecture Notes in Computer Science), Springer International Publishing, Cham, 594–613. DOI:https://doi.org/10.1007/978-3-319-58640-3_43

[9] Lei Cui, Shaohan Huang, Furu Wei, Chuanqi Tan, Chaoqun Duan, and Ming Zhou. 2017. SuperAgent: A Customer Service Chatbot for E-commerce Websites. In *Proceedings of ACL 2017, System Demonstrations*, Association for Computational Linguistics, Vancouver, Canada, 97–102.

[10] Barry E. Cushing and Sunita S. Ahlawat. 1996. Mitigation of recency bias in audit judgment: The effect of documentation. *Auditing* 15, 2 (Fall 1996), 110.

[11] Min-Yuh Day, Jian-Ting Lin, and Yuan-Chih Chen. 2018. Artificial intelligence for conversational robo-advisor. In *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE Press, 1057–1064.

[12] Smit Desai and Jessie Chin. 2021. Hey Google, Can You Help Me Learn? In *CUI 2021 - 3rd Conference on Conversational User Interfaces*, ACM, Bilbao (online) Spain, 1–4. DOI:https://doi.org/10.1145/3469595.3469601

[13] Diogo Ferreira, Filipe Portela, and Manuel Filipe Santos. 2021. A Step Towards the Use of Chatbots to Support the Enterprise Decision-Making Processes. In *Trends and Applications in Information Systems and Technologies* (Advances in Intelligent Systems and Computing), Springer International Publishing, Cham, 308–317. DOI:https://doi.org/10.1007/978-3-030-72654-6_30

[14] Baruch Fischhoff and Ruth Beyth. 1975. I knew it would happen: Remembered probabilities of once—future things. *Organizational Behavior and Human Performance* 13, 1 (February 1975), 1–16. DOI:https://doi.org/10.1016/0030-5073(75)90002-1

[15] Arthur C. Graesser. 2016. Conversations with AutoTutor Help Students Learn. *Int J Artif Intell Educ* 26, 1 (March 2016), 124–132. DOI:https://doi.org/10.1007/s40593-015-0086-4

[16] Arthur C. Graesser, Haiying Li, and Carol Forsyth. 2014. Learning by Communicating in Natural Language With Conversational Agents. *Curr Dir Psychol Sci* 23, 5 (October 2014), 374–380. DOI:https://doi.org/10.1177/0963721414540680

[17] Xu Han, Michelle Zhou, Matthew J. Turner, and Tom Yeh. 2021. Designing Effective Interview Chatbots: Automatic Chatbot Profiling and Design Suggestion Generation for Chatbot Debugging. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–15. DOI:https://doi.org/10.1145/3411764.3445569

[18] Thorsten Hens and Anna Meier. 2015. Behavioral finance: the psychology of investing. *Credit Suisse* (2015).

[19] Christian Hildebrand and Anouk Bergner. 2020. Conversational robo advisors as surrogates of trust: onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science* (November 2020). DOI:https://doi.org/10.1007/s11747-020-00753-z

[20] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03), Association for Computing Machinery, New York, NY, USA, 17–24. DOI:https://doi.org/10.1145/642611.642616

[21] Angel Hsing-Chi Hwang and Andrea Stevenson Won. 2021. IdeaBot: Investigating Social Facilitation in Human-Machine Team Creativity. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI '21), Association for Computing Machinery, New York, NY, USA, 1–16. DOI:https://doi.org/10.1145/3411764.3445270

[22] Soomin Kim, Jinsu Eun, Changhoon Oh, Bongwon Suh, and Joonhwan Lee. 2020. Bot in the Bunch: Facilitating Group Chat Discussion by Improving Efficiency and Participation with a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376785

[23] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM, Glasgow Scotland Uk, 1–13. DOI:https://doi.org/10.1145/3290605.3300932

[24] Yi-Chieh Lee, Naomi Yamashita, and Yun Huang. 2020. Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW1 (May 2020), 1–27. DOI:https://doi.org/10.1145/3392836

[25] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–12. DOI:https://doi.org/10.1145/3313831.3376175

[26] Lijia Lin, Paul Ginns, Tianhui Wang, and Peilin Zhang. 2020. Using a pedagogical agent to deliver conversational style instruction: What benefits can you obtain? *Computers & Education* 143, (January 2020), 103658. DOI:https://doi.org/10.1016/j.compedu.2019.103658

[27] Raju Maharjan, Darius Adam Rohani, Per Bækgaard, Jakob Bardram, and Kevin Doherty. 2021. Can we talk? Design Implications for the Questionnaire-Driven Self-Report of Health and Wellbeing via Conversational Agent. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*, ACM, Bilbao (online) Spain, 1–11. DOI:https://doi.org/10.1145/3469595.3469600

[28] Stefan Morana, Ulrich Gnewuch, Dominik Jung, and Carsten Granig. 2020. The Effect of Anthropomorphism on Investment Decision-Making with Robo-Advisor Chatbots. *ECIS 2020 Research Papers* (June 2020). Retrieved from https://aisel.aisnet.org/ecis2020_rp/63

[29] Jaya Narain, Tina Quach, Monique Davey, Hae Won Park, Cynthia Breazeal, and Rosalind Picard. 2020. Promoting Wellbeing with Sunny, a Chatbot that Facilitates Positive Messages within Social Groups. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–8. DOI:https://doi.org/10.1145/3334480.3383062

[30] Eben Otuteye and Mohammad Siddiquee. 2015. Overcoming Cognitive Biases: A Heuristic for Making Value Investing Decisions. *Journal of Behavioral Finance* 16, 2 (April 2015), 140–149. DOI:https://doi.org/10.1080/15427560.2015.1034859

[31] Hyanghee Park and Joonhwan Lee. 2021. Designing a Conversational Agent for Sexual Assault Survivors: Defining Burden of Self-Disclosure and Envisioning Survivor-Centered Solutions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–17. DOI:https://doi.org/10.1145/3411764.3445133

[32] SoHyun Park, Anja Thieme, Jeongyun Han, Sungwoo Lee, Wonjong Rhee, and Bongwon Suh. 2021. "I wrote as if I were telling a story to someone I knew.": Designing Chatbot Interactions for Expressive Writing in Mental Health. In *Designing Interactive Systems Conference 2021*, ACM, Virtual Event USA, 926–941. DOI:https://doi.org/10.1145/3461778.3462143

[33] Judith J. Prochaska, Erin A. Vogel, Amy Chieng, Matthew Kendra, Michael Baiocchi, Sarah Pajarito, and Athena Robinson. 2021. A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study. *Journal of Medical Internet Research* 23, 3 (March 2021), e24850. DOI:https://doi.org/10.2196/24850

[34] Rifat Rahman, Md. Rishadur Rahman, Nafis Irtiza Tripto, Mohammed Eunus Ali, Sajid Hasan Apon, and Rifat Shahriyar. 2021. AdolescentBot: Understanding Opportunities for Chatbots in Combating Adolescent Sexual and Reproductive Health Problems in Bangladesh. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, Yokohama Japan, 1–15. DOI:https://doi.org/10.1145/3411764.3445694

[35] Aditi Ramachandran, Chien-Ming Huang, Edward Gartland, and Brian Scassellati. 2018. Thinking Aloud with a Tutoring Robot to Enhance Learning. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '18), Association for Computing Machinery, New York, NY, USA, 59–68. DOI:https://doi.org/10.1145/3171221.3171250

[36] Neal J. Roese and Kathleen D. Vohs. 2012. Hindsight Bias. *Perspect Psychol Sci* 7, 5 (September 2012), 411–426. DOI:https://doi.org/10.1177/1745691612454303

[37] Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L. Murnane, Emma Brunskill, and James A. Landay. 2019. QuizBot: A Dialogue-based Adaptive Learning System for Factual Knowledge. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), Association for Computing Machinery, Glasgow, Scotland Uk, 1–13. DOI:https://doi.org/10.1145/3290605.3300587

[38] Pavel Smutny and Petra Schreiberova. 2020. Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education* 151, (July 2020), 103862. DOI:https://doi.org/10.1016/j.compedu.2020.103862

[39] Donggil Song, Eun Young Oh, and Marilyn Rice. 2017. Interacting with a conversational agent system for educational purposes in online courses. In *2017 10th International Conference on Human System Interactions (HSI)*, 78–82. DOI:https://doi.org/10.1109/HSI.2017.8005002

[40] Vivian Ta, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero, and Alexia Loggarakis. 2020. User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. *Journal of Medical Internet Research* 22, 3 (March 2020), e16235. DOI:https://doi.org/10.2196/16235

[41] Yin-Te Tsai and Wei-An Lin. 2018. Design of an Intelligent Cognition Assistant for People with Cognitive Impairment. In *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 1207–1212. DOI:https://doi.org/10.1109/HPCC/SmartCity/DSS.2018.00203

[42] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka, Matcheri S. Kashavan, and John Blake Torous. 2019. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Can J Psychiatry* 64, 7 (July 2019), 456–464. DOI:https://doi.org/10.1177/0706743719828977

[43] Thiemo Wambsganss, Tobias Kueng, Matthias Soellner, and Jan Marco Leimeister. 2021. ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (CHI '21), Association for Computing Machinery, New York, NY, USA, 1–13. DOI:https://doi.org/10.1145/3411764.3445781

[44] Martin Weber and Colin F. Camerer. 1998. The disposition effect in securities trading: an experimental analysis. *Journal of Economic Behavior & Organization* 33, 2 (January 1998), 167–184. DOI:https://doi.org/10.1016/S0167-2681(97)00089-9

[45] Rainer Winkler, Sebastian Hobert, Antti Salovaara, Matthias Söllner, and Jan Marco Leimeister. 2020. Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), Association for Computing Machinery, Honolulu, HI, USA, 1–14. DOI:https://doi.org/10.1145/3313831.3376781

[46] Rainer Winkler, Matthias Söllner, Maya Lisa Neuweiler, Flavia Conti Rossini, and Jan Marco Leimeister. 2019. Alexa, Can You Help Us Solve This Problem?: How Conversations With Smart Personal Assistant Tutors Increase Task Group Outcomes. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA '19*, ACM Press, Glasgow, Scotland Uk, 1–6. DOI:https://doi.org/10.1145/3290607.3313090

[47] Ziang Xiao, Michelle X. Zhou, Wenxi Chen, Huahai Yang, and Changyan Chi. 2020. If I Hear You Correctly: Building and Evaluating Interview Chatbots with Active Listening Skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ACM, Honolulu HI USA, 1–14. DOI:https://doi.org/10.1145/3313831.3376131

[48] Ziang Xiao, Michelle X. Zhou, Q. Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-ended Questions. *ACM Trans. Comput.-Hum. Interact.* 27, 3 (June 2020), 15:1-15:37. DOI:https://doi.org/10.1145/3381804

[49] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. 2017. A New Chatbot for Customer Service on Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 3506–3510.

[50] Yingzi Xu, Chih-Hui Shieh, Patrick van Esch, and I-Ling Ling. 2020. AI customer service: Task complexity, problem-solving ability, and usage intention. *Australasian Marketing Journal (AMJ)* 28, 4 (November 2020), 189–199. DOI:https://doi.org/10.1016/j.ausmj.2020.03.005

[51] Syed Aliya Zahera and Rohit Bansal. 2018. Do investors exhibit behavioral biases in investment decision making? A systematic review. *QRFM* 10, 2 (May 2018), 210–251. DOI:https://doi.org/10.1108/QRFM-04-2017-0028