

Planning Habit: Daily Planning Prompts with Alexa ^{*}

Andrea Cuadra^[0000-0002-1845-0240], Oluseye Bankole, and
Michael Sobolev^[0000-0002-8931-7682]

Cornell Tech, New York, NY, 10044, USA

Abstract. The widespread adoption of intelligent voice assistants (IVAs), like Amazon’s Alexa or Google’s Assistant, presents new opportunities for designers of persuasive technologies to explore how to support people’s behavior change goals and habits with voice technology. In this work, we explore how to use *planning prompts*, a technique from behavior science to make specific and effective plans, with IVAs. We design and conduct usability testing ($N=13$) on a voice app called *Planning Habit* that encourages users to formulate daily plans out loud. We identify strategies that make it possible to successfully adapt planning prompts to voice format. We then conduct a week-long online deployment ($N=40$) of the voice app in the context of daily productivity. Overall, we find that traditional forms of planning prompts can be adapted to and enhanced by IVA technology.

1 Introduction

People encounter problems in translating their goals into action [16]—often termed the intention-behavior gap. *Planning prompts* is a technique that can help people make concrete and specific plans that they are more likely to follow-through on than larger, less achievable goals. Planning prompts have been demonstrated to be an effective, self-regulatory strategy in domains such as flu vaccination, voting, and insurance [39]. Research in behavior change and persuasive technology has begun to explore the implementation of planning prompts for habit formation [33]. There is an opportunity to expand the use planning prompts to, now mainstream, IVAs.

IVAs have unique potential for persuasive design, because they can be used in an eyes- and hands-free manner and can be more intuitive to use for non-digital natives [35]. *We envision IVAs as a useful platform for transitioning planning prompts to voice format*, thereby expanding the opportunities for IVAs to support positive behavior change. However, the design of these interactions is complex [17], and thus requires careful attention and iteration. We present an exploratory study that examines how to adapt planning prompts from written to voice format (Figure 1).

^{*} We are grateful for the support received from Deborah Estrin, Nicola Dell, and our voluntary participants. This work was supported by the National Science Foundation under Grant No. 1700832.



Fig. 1. Example interaction with *Planning Habit*, the Amazon Alexa skill, or voice app, created for this study. A user asks Alexa to open the *Planning Habit* skill. Alexa responds by stating the user's goal and instructing the participant to focus on one plan that will help her achieve that goal.

We make two contributions to research on persuasive technology and behavior change systems. First, we present finding from a research through design approach [49] for adapting planning prompts to IVA interactions. We design a voice app called *Planning Habit* that encourages users to formulate daily plans out loud. The design process surfaced common and generalizable challenges and allowed use to develop strategies to overcome these challenges. Second, we provide evidence for the use of IVAs to elicit spoken planning prompts from our quantitative and qualitative findings from a week-long feasibility study deployed via Amazon Mechanical Turk (mTurk), $N=40$. These contributions are a unique result of the mixed methods employed: iterative design, in-the-wild deployment, and qualitative and quantitative analysis, and will be useful for researchers interested in designing IVA in behavior change systems and persuasive technologies.

2 Related Work

IVAs are voice-based software agents that can perform tasks upon request—they use natural language processing to derive intent from requests made by their users, and respond to those requests using speech and/or another modality (e.g., graphical output) [42]. We focused on the intersections of IVAs and behavior change which is currently nascent. We discuss existing work surrounding IVAs for behavior change and planning prompts research in context of persuasive and behavior change technology.

2.1 IVAs

Multiple lines of research recognize the potential of IVAs, and new research is emerging in many areas. One line of work focuses on technical advances, including distant speech recognition [21], human-sounding text-to-speech [24], and

question answering, natural language inference, sentiment analysis, and document ranking [11] [46]. Another line of work focuses on risks IVAs may introduce, including privacy concerns [22] [8], vulnerabilities to attackers [28] [8] [48], inconsistent and incomplete answers to simple questions about mental health [27], and possible pitfalls that may occur in medical settings, such as misrecognition of medical names, or unexpected input [4]. A third line of research looks at IVAs at a more meta-level, characterizing current use and impact by analyzing usage logs [3], identifying trends from product reviews [36] [30], or comparing different commercial IVAs [27] [25] [37]. Researchers have also examined the social role of IVAs [36] [6] [7], their conversational (or not) nature [2] [9], their ability to help young children read and learn [45], and their promise as a tool to encourage self-disclosure [23] [47].

Work at the intersections of IVAs and behavior change is more nascent. One example is “FitChat”, which was developed by Wiratunga, et al. to encourage physical activity among older adults [44]. This study found that voice is a powerful mode of delivering effective digital behavior change interventions, which may increase adherence to physical activity regimes and provide motivation for trying new activities [44]. Sezgin et al. provide a scoping review of patient-facing, behavioral health interventions with voice assistant technologies that target self-management and healthy lifestyle behaviors [41]. However, this scoping review also includes many research papers using interactive voice response (IVR) systems [41], which are different from IVAs (we consider IVR systems to be the less-capable, usually telephone-based predecessors to IVAs). The study found that voice assistant technology was generally used to either: a) deliver education or counseling/skills, or b) monitor/track self-reported behaviors. It also found that research-adapted voice assistants, in contrast to standalone commercially available voice assistants, performed better regarding feasibility, usability, and preliminary efficacy, along with high user satisfaction, suggesting a role for voice assistants in behavioral health intervention research [41]. Our research explores a new perspective to the literature on IVAs and behavior change by examining how to adapt planning prompts, a behavior science technique, from written to voice format.

2.2 Planning Prompts

Planning prompts are a simple and effective behavioral technique to translate their goals into action [39]. Gollwitzer famously argued that simple plans can have a strong effect on goal achievement [16]. Planning prompts are subordinate to goals and specify “when, where and how” goals might be achieved while goals themselves specify “what” needs to be achieved [16]. Plans can be considered planning prompts if they contain specific details as described above. In a recent review, planning prompts were argued to be simple, inexpensive, and powerful nudges that help people do what they intend to get done [39]. Prior research has explored integrating implementation intentions into mobile devices by using contextual triggers and reinforcement was explored as a mechanism for habit formation [43] [34]. In the context of digital well-being, the Socialize Android

app [38] was developed with user-specified implementation intentions to replace undesired phone usage with other desired activities or goals. The process of generating action plans can be partly or completely automated, as exemplified by TaskGenies [20]. In the context of physical activity, DayActivizer [12] is a mobile app that tries to encourage physical activity by generating plans from contextual activity data. Contextual factors such as previous activity, location and time can help generate better plans for individuals [32]. A recent review of digital behavior change also highlighted the potential of implementation intentions for habit formation [33]. Because of the potential that IVAs may have to encourage behavior change, it is imperative that more research is conducted in this topic.

3 Design Process

In this work, we employ a research through design approach [49] [18] [13] to explore how the behavioral science technique of planning prompts might be adapted from written to spoken format. Our design goal was to create a voice app or *skill* (Amazon’s terminology for apps that run on their Alexa platform) that elicits spoken-out-loud planning prompts (see Figure 1). We relied on evidence-based techniques from behavioral science paired with an iterative design process to make this technology engaging and persuasive. We now describe the three stages of our design process: our initial design, usability testing, and the final design.

3.1 Stage I: Initial Design

Our initial design is grounded in previous research on planning prompts for behavior change and habit formation [16] [39] and persuasive and behavior change technologies [14] [10]. Drawing on insights from this research, we formulated the following initial guidelines to ground our first prototype before evaluating it via user testing:

1. **Behavior science suggests that planning prompts will work aloud:** A *planning prompt’s* purpose is to nudge people to think through how and when they will follow through with their plans [39] [16]. Although the literature about planning prompts predominantly uses examples about written prompts [39] [16], voice planning prompts may fulfill the same purpose. Thus, we formulated our first assumption—that planning prompts will also be effective at fulfilling the same purpose if they are spoken aloud.
2. **HCI research tells us users will need the voice app to allow them to interact with their plans:** Consolvo et al. in her guidelines for behavior change technology highlight the need for technology to be controllable [10]. In natural settings, people have the ability to revisit, revise, and “check-off” our plans, especially when written down. Thus, we planned for our digital skill to mimic those affordances by allowing users to interact with their plans.

We relied on these guidelines to inform our initial design. When a user opened the voice app, it asked the user whether or not she had completed the previous plan. If affirmative, the voice app gave the user the option to make a new plan. Otherwise, the user was given the opportunity to keep the previous plan.

Implementation We built the voice app using Amazon Alexa, because of its popularity and robust developer ecosystem. We used Alexa Skills Kit (ASK), which is a compilation of open sourced Alexa application programming interfaces and tools to develop voice apps. We stored usage data in an external database.

3.2 Stage II: Usability Testing

Our initial research plan incorporated the Rapid Iterative Testing and Evaluation method (RITE method) [26] to ensure that our skill was easy to use. The RITE method is similar to traditional usability testing, but it advocates that changes to the user interface are made as soon as a problem is identified and a solution is clear [26]. We conducted usability tests (N=13) with university students. At the beginning tests were performed in a lab setting (N=10). Subsequently, usability testing was conducted in participant’s apartments (N=3).

For initial usability testing, participants were asked to create plans over a period of three simulated days, and then tell us about their experience using the skill. Each usability test lasted about 15 minutes. We spread out the usability tests over two weeks to allow for time to make design adjustments based on findings from these sessions. This testing exposed major challenges with the technology’s transcription accuracy:

1. **The name of the skill was frequently misheard:** the name of the skill was originally “Planify.” In usability tests, we found that Alexa did not recognize the invocation phrase, “open Plan-ify”, when the participant did not have an American accent. Instead, it would frequently suggest opening other skills with the word “planet”.
2. **The plans were incorrectly transcribed:** plans often had specific keywords that were misrecognized. For example, “*NSF proposal*” was transcribed to “NBA proposal,” completely changing the meaning of the plan. This created confusion in two parts of our skill design: 1) when the IVA repeated the previous day’s plan to the user, and 2) when the IVA confirmed the new plan.

We redesigned the skill to work around these challenges. We renamed the skill “Planning Habit,” which was easier to consistently transcribe across accents. We also redesigned the skill so that it would not have to restate (nor understand) the plan after the participant said it, which was counter HCI guidelines surrounding giving visibility of the system’s status and control [10] [29]. This was a deliberate effort needed to overcome limitations inherent to current language recognition technologies. The resulting interaction only had three steps: 1) request a plan,

2) listen to plan, and 3) end session by requesting the participant to check-in again the next day.

Once usability errors in the lab setting became rare, we conducted usability testing in participants' own homes to: 1) test that the Amazon Alexa skill we had developed worked over an extended period of time in participants' homes, and 2) test the back-end tracking of participants' daily interactions with *Planning Habit*. We stored interaction data in an external database. The data stored included the transcripts of the voice snippets of plans, the timestamps of each interaction, and the associated user identification number. This data helped us understand how each participant was interacting with the skills, and probe deeper when interviewing them about their experience later on. We recruited university students participants who already owned Amazon Alexa devices, were willing to install the skill, use it every day for at least a week, and participate in a 30-minute interview at the end of the study. We did not offer compensation. We gave participants minimal instructions—to say, “Alexa, open Planning Habit,” and then make a plan that would help them be more productive every day. For each interview two researchers were present, one asked questions and the other took notes. We asked participants to tell us about their experience, to describe the sorts of plans they made, how (if at all) the skill had affected them, what worked well, and what worked poorly. After each question, we dove deeper by asking for more details. For example, if a participant mentioned they stopped using the skill, we would ask why. For this part of the study, all participants used an Alexa smart speaker.

During the at-home testing phase, we saw glimpses of both how the Planning Habit tool might benefit participants along with further limitations of the tool. The benefits included:

- **Accountability.** One participant said that the skill affected him, because “*when [he] said [he] would do it, then [he] would.*”
- **Help with prioritization.** Participants surfaced the skill’s role in helping them prioritize, “*it’s a good thing to put into your morning routine, if you can follow along with it it’s a good way to plan your day better and think about what you have to prioritize.*”
- **Ease of use.** Participants commented on the ease, “*it’s easy to incite an Alexa, and easy to complete the [planning] task.*”
- **Spoken format leading to more complete thoughts.** One participant said, “*it sounds weird when you say these things aloud, in that it feels like a more complete thought by being a complete sentence, as opposed to random tidbits of things.*”

The limitations included:

- **Difficulty remembering the command and making effective plans.** Participants commented that it was difficult to remember the command to make the plans, and suggested that “*it would be useful to remind people of the command on every communication.*”

- **Making effective plans.** Participants indicated that they did not have enough guidance about how to make their plans, or what sorts of plans to make. This need corresponds to previous research in behavioral science that highlights the need for training to use planning prompts effectively [33].
- **Error proneness.** Participants commented on the skill being “*very error prone.*” Many of these errors had to do with Alexa abruptly quitting the skill for unknown reasons, or because the user paused to think mid-plan. Alexa comes configured to listen for at most 8 seconds of silence, and Amazon does not give developers the ability to directly change that configuration. A participant stated, “*a couple of times I was still talking when it closed its listening period, and that made me think that ‘huh, maybe Alexa is not listening to me right now.’*” Formulating planning prompts on the spot can require additional time to consider possible options, and make a decision about which one to pick.
- **Limited meaningfulness.** One participant indicated that he did not observe any meaningful improvement in how easy his life felt after having used the skill saying, “*I don’t think it made my life any easier or anything of that nature.*” Furthermore, many of the plans made, as judged by authors with behavior science expertise, were not likely to be effective. This suggests that participants were also not experiencing the benefits of getting closer to attaining a goal.

We explain how we addressed these issues in Section 3.3.

3.3 Stage III: Final Design

Based on the findings from the testing stage we restructured the skill to follow a structure that would provide more guidance and more motivation, and avoid transcription errors. We structured the conversation using the following components:

1. A **greeting** to create a personable start.
2. A **rationale** to increase motivation and understanding.
3. The **participant’s goal** to establish grounding for making plans related to a goal the participant wants to attain, and thus to improve ability and motivation. We asked participants to type three work-related goals in an onboarding questionnaire. We then personalized each participant’s experience by copy-pasting their responses to the voice app’s script.
4. A **planning tip** to improve ability to create effective plans.
5. **Thinking time** to allot extra time to formulate a plan.
6. A **goodbye tip** to improve ability to follow-through with their plans.

Additionally, we asked participants to include the daily command in their reminder, in order to reduce difficulty remembering the command. Each participant was instructed tell their Alexa device, “Alexa, set a daily reminder to open Planning Habit at [time in the morning].” We also added “thinking time”, by playing background music for 14 seconds (with the option to ask for more when

they expired) to give users extra time to think about their plans. By adding the music we were able to set clear expectations for the interaction, and avoid having Alexa quit before the user was ready to say their plan.

The final design had fewer steps than the original one, and included guidance after the plan is made. We selected a combination of different options for each component of the formula (i.e., greeting, rationale, participant’s goal, planning tip, thinking time, and goodbye tip), and rotated them each day of the study.

4 Feasibility Study

We conducted a feasibility study on MTurk to explore the effects of Planning Habit with participants in more natural setting. To do so, we built on the work of Okeke et al., who previously used a similar approach deploying interventions to MTurk participants’ existing devices [31]. Our goals were to understand what sorts of plans people would make, engagement with the voice app, and their opinions surrounding satisfaction, planning behavior improvement, and overall strengths and weaknesses of the voice app.

4.1 Method

We deployed the voice app for a period of one week with 40 mTurk participants. We asked participants to complete an on-boarding questionnaire, and instructed participants to install the skill and set daily reminders. Then, we asked participants to make a plan using the skill every day for six days. Last, we asked participants to fill out a post-completion questionnaire of the skill at the end of study. All study procedures were exempted from review by Cornell University’s IRB under Protocol 1902008577.

Participants and Procedure A total of $N=40$ participants (18F, 22M) passed all the checks we put in place. These checks included trick questions, submission of a screenshot of the daily reminder set on an Alexa device, a skill installation identification number, and back-end verification of usage logs. All participants indicated they interacted with Alexa (broadly, not specifically with *Planning Habit*) at least once a week, and most said they used it daily ($N=25$). Most participants indicated they owned multiple Alexa devices ($N=22$).

Participants were instructed to go to the Alexa skill store and install “Planning Habit”. Then, they had to open the skill on their device, and enter the ID number that the app gave them into a questionnaire. Then, they were asked a series of demographic, and Alexa-usage questions. Next, they had to write three work-related goals (which were then incorporated into each participant’s personalized voice app). Finally, participants were asked to set a reminder, and upload screenshot as evidence.

Participants were compensated \$5 for completing the on-boarding questionnaire and installing the skill, and given a \$10 bonus for interacting with the skill throughout the week and completing the final questionnaire. All $N=40$

participants who successfully installed the skill and completed the on-boarding questionnaire received \$5. Only 22 participants were eligible to receive full participation bonus of \$10 at the end of the study. A few participants ($N=2$) that demonstrated reasonable engagement, but did not fulfill all the requirements, received a reduced bonus of \$5.

Measures and Evaluation We evaluated the feasibility of adapting planning prompts from written to voice format by qualitatively analyzing usage data alongside responses from the post-completion subjective evaluations. For usage data, we measured the number of times each participant made a plan using *Planning Habit*, and searched for patterns or interesting insights in the plans they created. For subjective evaluation, we asked participants about their satisfaction with the voice app, self-perception of improvement in planning ability, and likeliness to continue using the skill or recommend to it to others.

4.2 Quantitative Findings

Most participants made at least 3 plans throughout the duration of the study. Engagement results are based on the participants that successfully completed the on-boarding questionnaire and skill installation ($N=40$). The metadata revealed that more than a third of the participants ($N=14$) demonstrated 100% engagement, completing 6/6 plans.¹ Several participants ($N=11$) made between 3 and 5 plans in total. A few participants ($N=4$) made between 1 and 2 plans. The rest of the participants ($N=11$) never made any plans.

For the rest of the sections, we report findings based on only the participants ($N=22$) that interacted with the skill at least 3 days during the intervention, and completed the final questionnaire. We excluded responses from participants that did not complete more than 2 plans, as this level of usage does not constitute sufficient engagement with the skill to provide a genuine assessment. The discarded responses were extremely positive and vague. We ended up with a total of 129 plans for analysis after discarding plans from participants that did not sufficiently engage with the voice app.

Most participants were somewhat or extremely satisfied with the skill. Most (77%) reported that they were somewhat or extremely satisfied, some participants (18%) felt neutral, and a few (5%) reported dissatisfaction with the skill. Furthermore, when asked whether they would recommend the skill to others, most participants (59%) indicated they were at least somewhat likely to recommend to a close family member or friend. In addition, some participants (32%) said they would continue using the skill and only a few (14%) said they would not; the remaining 54% were unsure.

¹ There are 6 total plans, and not 7, because the first day of the week-long study was reserved for installation of the skill.

Most participants indicated the skill helped them become better planners. Most participants (59%) indicated the skill helped them become better planners overall, suggesting that *Planning Habit* may be a feasible way to improve people’s ability to make effective plans.

4.3 Qualitative Findings

We analyzed our qualitative data (the plans participants made throughout the deployment, and the open-ended questionnaire responses) by organizing the data based on observed plan components, and comments surrounding satisfaction. The authors individually categorized the data, and met to discuss and reconcile differences.

1. Participants’ plans incorporated the tips for effective planning we provided via our voice app. They did so by:

- **Indicating a location, a time, or a way of completing their plan.** For example, this plan mentions locations and a time, “*planet [sic] taking my kids to school and then going to the gym.*”² The locations include school, and the gym. The time is relative, after taking the kids to school. Another participant made a plan to “*analyze at least 5 different distributions,*” in which the participant specified specific details (5 different distributions) about a way to complete the plan. Per our categorization of the plans, 75% indicated a location, a time, or a way of completing their plan, 16% did not, and 9% were too ambiguous for us to categorize.
- **Making plans centered around their bigger goals.** A participant whose goal was to “*do more daily ed activities with [her] daughter*” made a plan to “*take [her] daughter to story time at [their] local library,*” a daily ed activity. We counted the number of plans that related to the participant’s goals, and we found that 59% related to the goals, 11% did not relate to the goals, and 30% of plans were too ambiguous for us to determine whether they related to a goal or not.
- **Thinking about the things in the way of their goals, and how to overcome those obstacles.** One of the Planning Tips we provided said, “Take a moment to think about the things in the way of your goal. What’s a single task that will help you overcome one of these?” One participated reacted to the tip to think about an obstacle by uttering the obstacle, “*[first] of all I don’t have any money to renovate that trailer.*” Another made a plan and provided an explanation of how his plan would help him overcome a scheduling obstacle, “*book a meeting for 4:30 so I can get out a little bit earlier today and go me[et] the client*”. Our counts revealed that 19% of the

² In this plan, we may also observe the type of transcription inaccuracies that were surfaced earlier in the study—the word “plan” plus some other utterance was transcribed to “planet”. In this part of the study, the transcriptions were available to us in the back-end, but not surfaced to the user.

plans mentioned an obstacle of some kind, 73% did not, and 8% were too ambiguous for us to categorize.

2. Participants valued the voice app’s guidance, but wanted to track their plans. Participants found the guidance from the skill to be valuable, and frequently mentioned that using the skill helped them think about their daily priorities and plan accordingly. Many responses echoed these sentiments, e.g., “*it was a good skill to get you to stop and think and plan out actions,*” or “*it was helpful to know what my priority was for the day.*” However, participants continued to express the need to track plan completion. Some participants indicated feeling limited by the lack of follow-through. For example, one participant said, “*I like the idea of starting the day with a plan but it almost feels like an empty gesture with no consequences or follow-up.*” This constraint could potentially be solved if the skill was able to accurately transcribe people’s plans and remind them what they said, but as described in Section 3.2, the inaccuracy of transcription hampered our ability to implement such tracking.

5 Discussion

We demonstrate the feasibility of adapting the behavioral technique of planning prompts from text to voice. Our planning prompt voice app proved to be easy-to-use and effective, which serves to validate the initial development work we did. Together, the incorporation of our tips in participants’ plans, the relatively high levels of engagement with the voice app and satisfaction, and participants’ perceived improvement planning behavior, suggest that traditional forms of planning prompts can be adapted to and enhanced by IVA technology.

We encountered several challenges with the state-of-the-art of voice technologies that will be mitigated as the technology continues to improve. Many speech recognition milestones—such being able to recognize different voices, speech at different speeds, from different locations in a room—had to be achieved to let us interact with IVAs the way we do today [15]. There are many efforts to continue improving these technologies. For example, Mozilla’s Common Voice dataset is part of an effort to bridge the digital speech divide, allowing people all over the world to contribute to it and to download the dataset to train speech-enabled applications [1]. Doing so, will allow the technology to become better at recognizing more people’s speech, a problem we encountered during our initial usability sessions, as described in Section 3.2. Similarly, although the ability to track changes is considered an important aspect of both behavior change technologies [10], the technology was not sufficiently advanced to be able to do so effectively. However, speech recognition technology is actively improving [19] [40], meaning these challenges will eventually disappear. Currently, the technological constraints we encountered may hinder engagement, so it is important to continue considering features such as plan-tracking as the technology’s ability to understand speech improves.

The promise of voice technology extends beyond speech recognition improvements. For example, understanding specific contexts for generating guidance can generate immense value. When guiding users to create effective planning prompts, it is important not to only transcribe the speech, but also understand the content in the plan (e.g., the plan’s associated goal, when and where the plan is scheduled to happen, etc.), and to appropriately schedule the timing of the reminder. Using automation to understand the content in a plan could help generate personalized guidance to maximize a person’s ability to create effective planning prompts. Furthermore, Cha et al. are generating research surrounding opportune moments for proactive interactions with IVAs, and identifying contextual factors, such as resource conflicts or user mobility, that may play an important role in interactions initiated by IVAs [5]. Such advancements could mean that we could design reminders to happen not just at a set time, but at opportune moments.

5.1 Limitations and Future Research

The exploratory nature of the study comes with its limitations. When we interacted with participants in person during our design process, we were able to understand nuances of the interactions in depth. Doing so allowed us to evolve the design of the voice app to the one we used for the feasibility study. However, during the feasibility study, we collected data automatically via questionnaires and usage logs, and did not have the opportunity to ask participants questions in real-time. By studying the voice app in a less-controlled setting, we were able to observe that many participants were highly engaged and found the voice app helpful. However, a hands-off deployment can introduce bias when subjectively classifying the plans participants made, because researchers cannot confirm their judgments with study participants. In our case, the inability to consult with participants during the feasibility study also added noise to our data, since we had to classify many plans as “other” due to ambiguity, or missing information. Finally, due to its exploratory nature, a long-term evaluation was outside of scope. Despite the limitations of this work, our design process and feasibility study allowed us to create a detailed picture of participants’ experience using the our voice app, and generate valuable contributions.

6 Conclusion

This paper contributes a design exploration of implementing planning prompts, a concept for making effective plans from behavior science, using IVAs. We found that traditional forms of planning prompts can be adapted to and enhanced by IVA technology. We surfaced affordances and challenges specific to IVAs for this purpose. Finally, we validated the promise of our final design through an online feasibility study. Our contributions will be useful for improving the state-of-the-art of digital tools for planning, and provide insights for others interested in adapting insights and techniques from behavior science to interactions with IVAs.

References

1. Common voice by mozilla. Retrieved 2020-03-03 from <https://voice.mozilla.org/en/datasets>.
2. BENETEAU, E., RICHARDS, O. K., ZHANG, M., KIENZ, J. A., YIP, J., AND HINIKER, A. Communication breakdowns between families and alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019), pp. 1–13.
3. BENTLEY, F., LUVOGT, C., SILVERMAN, M., WIRASINGHE, R., WHITE, B., AND LOTTRIDGE, D. Understanding the long-term use of smart speaker assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–24.
4. BICKMORE, T., TRINH, H., ASADI, R., AND OLAFSSON, S. Safety first: Conversational agents for health care. In *Studies in Conversational UX Design*. Springer, 2018, pp. 33–57.
5. CHA, N., KIM, A., PARK, C. Y., KANG, S., PARK, M., LEE, J., LEE, S., AND LEE, U. “hello there! is now a good time to talk?”: Opportune moments for proactive interactions with smart speakers.
6. CHA, Y., HONG, Y., JANG, J., AND YI, M. Y. “Jack-of-all-trades”: A thematic analysis of conversational agents in multi-device collaboration contexts. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–6.
7. CHIN, H., AND YI, M. Y. Should an agent be ignoring it? A study of verbal abuse types and conversational agents’ response styles. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–6.
8. CHUNG, H., IORGA, M., VOAS, J., AND LEE, S. Alexa, can i trust you? *Computer* 50, 9 (2017), 100–104.
9. CLARK, L., PANTIDI, N., COONEY, O., DOYLE, P., GARAIALDE, D., EDWARDS, J., SPILLANE, B., GILMARTIN, E., MURAD, C., MUNTEANU, C., WADE, V., AND COWAN, B. R. What makes a good conversation? Challenges in designing truly conversational agents. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–12.
10. CONSOLVO, S., McDONALD, D. W., AND LANDAY, J. A. Theory-driven Design Strategies for Technologies That Support Behavior Change in Everyday Life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2009), CHI ’09, ACM, pp. 405–414. event-place: Boston, MA, USA.
11. DEVLIN, J., CHANG, M., LEE, K., AND TOUTANOVA, K. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR abs/1810.04805* (2018).
12. DOGANGÜN, A., SCHWARZ, M., KLOPPENBORG, K., AND LE, R. An approach to improve physical activity by generating individual implementation intentions. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (New York, NY, USA, 2017), UMAP ’17, ACM, pp. 370–375.
13. EASTERDAY, M. W., REES LEWIS, D. G., AND GERBER, E. M. The logic of design research. *Learning: Research and Practice* 4, 2 (2018), 131–160.
14. FOGG, B. J. Persuasive Technology: Using Computers to Change What We Think and Do. *Ubiquity 2002*, December (Dec. 2002).
15. GOLD, B., MORGAN, N., AND ELLIS, D. *Speech and audio signal processing: processing and perception of speech and music*. John Wiley & Sons, 2011.

16. GOLLWITZER, P. M. Implementation intentions: Strong effects of simple plans. *American Psychologist* 54, 7 (1999), 493–503.
17. GUZMAN, A. L. Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior* 90 (2019), 343–350.
18. HINIKER, A., HONG, S. R., KOHNO, T., AND KIENZT, J. A. Mytime: Designing and evaluating an intervention for smartphone non-use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016), CHI '16, ACM, pp. 4746–4757.
19. JIA, Y., JOHNSON, M., MACHEREY, W., WEISS, R. J., CAO, Y., CHIU, C.-C., ARI, N., LAURENZO, S., AND WU, Y. Leveraging weakly supervised data to improve end-to-end speech-to-text translation. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2019), IEEE, pp. 7180–7184.
20. KOKKALIS, N., KÖHN, T., HUEBNER, J., LEE, M., SCHULZE, F., AND KLEMMER, S. R. Taskgenies: Automatically providing action plans helps people complete tasks. *ACM Trans. Comput.-Hum. Interact.* 20, 5 (Nov. 2013), 27:1–27:25.
21. KUMATANI, K., MCDONOUGH, J., AND RAJ, B. Microphone array processing for distant speech recognition: From close-talking microphones to far-field sensors. *IEEE Signal Processing Magazine* 29, 6 (2012), 127–140.
22. LAU, J., ZIMMERMAN, B., AND SCHAUB, F. Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–31.
23. LEE, Y.-C., YAMASHITA, N., HUANG, Y., AND FU, W. "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* (2020), pp. 1–12.
24. LI, N., LIU, S., LIU, Y., ZHAO, S., LIU, M., AND ZHOU, M. Close to human quality TTS with transformer. *CoRR abs/1809.08895* (2018).
25. LÓPEZ, G., QUESADA, L., AND GUERRERO, L. A. Alexa vs. siri vs. cortana vs. google assistant: a comparison of speech-based natural user interfaces. In *International Conference on Applied Human Factors and Ergonomics* (2017), Springer, pp. 241–250.
26. MEDLOCK, M. C., WIXON, D., TERRANO, M., ROMERO, R., AND FULTON, B. Using the rite method to improve products: A definition and a case study. *Usability Professionals Association* (2002).
27. MINER, A., MILSTEIN, A., AND SCHUELLER, S. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health (vol 176, pg 619, 2016). *JAMA INTERNAL MEDICINE* 176, 5 (2016), 719–719.
28. MITEV, R., MIETTINEN, M., AND SADEGHI, A.-R. Alexa lied to me: Skill-based man-in-the-middle attacks on virtual assistants. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security* (2019), pp. 465–478.
29. NIELSEN, J. Ten usability heuristics, 2005.
30. O'BRIEN, K., LIGGETT, A., RAMIREZ-ZOHFELD, V., SUNKARA, P., AND LINDQUIST, L. A. Voice-controlled intelligent personal assistants to support aging in place. *Journal of the American Geriatrics Society* 68, 1 (2020), 176–179.
31. OKEKE, F., SOBOLEV, M., DELL, N., AND ESTRIN, D. Good Vibrations: Can a Digital Nudge Reduce Digital Overload? In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (New York, NY, USA, 2018), MobileHCI '18, ACM, pp. 4:1–4:12. event-place: Barcelona, Spain.

32. PARUTHI, G., RAJ, S., COLABIANCHI, N., KLASNJA, P., AND NEWMAN, M. W. Finding the sweet spot(s): Understanding context to support physical activity plans. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1 (Mar. 2018), 29:1–29:17.
33. PINDER, C., VERMEULEN, J., COWAN, B. R., AND BEALE, R. Digital Behaviour Change Interventions to Break and Form Habits. *ACM Transactions on Computer-Human Interaction* 25, 3 (June 2018), 15:1–15:66.
34. PINDER, C., VERMEULEN, J., WICAKSONO, A., BEALE, R., AND HENDLEY, R. J. If this, then habit: Exploring context-aware implementation intentions on smartphones. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (New York, NY, USA, 2016), MobileHCI '16, ACM, pp. 690–697.
35. PRADHAN, A., MEHTA, K., AND FINDLATER, L. "accessibility came by accident" use of voice-controlled intelligent personal assistants by people with disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), pp. 1–13.
36. PURINGTON, A., TAFT, J. G., SANNON, S., BAZAROVA, N. N., AND TAYLOR, S. H. "alexia is my new bff" social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (2017), pp. 2853–2859.
37. REIS, A., PAULINO, D., PAREDES, H., BARROSO, I., MONTEIRO, M. J., RODRIGUES, V., AND BARROSO, J. Using intelligent personal assistants to assist the elderly: an evaluation of amazon alexa, google assistant, microsoft cortana, and apple siri. In *2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)* (2018), IEEE, pp. 1–5.
38. ROFFARELLO, A. M., AND DE RUSSIS, L. Towards detecting and mitigating smartphone habits. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (New York, NY, USA, 2019), UbiComp/ISWC '19 Adjunct, ACM, pp. 149–152.
39. ROGERS, T., MILKMAN, K. L., JOHN, L. K., AND NORTON, M. I. Beyond good intentions: Prompting people to make plans improves follow-through on important tasks. *Behavioral Science & Policy* 1, 2 (2015), 33–41.
40. SARI, L., THOMAS, S., AND HASEGAWA-JOHNSON, M. Training spoken language understanding systems with non-parallel speech and text. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2020), IEEE, pp. 8109–8113.
41. SEZGIN, E., MILITELLO, L., HUANG, Y., AND LIN, S. A scoping review of patient-facing, behavioral health interventions with voice assistant technology targeting self-management and healthy lifestyle behaviors. *Behavioral Health Interventions with Voice Assistant Technology Targeting Self-management and Healthy Lifestyle Behaviors (April 1, 2019)* (2019).
42. SHUM, H.-Y., HE, X.-D., AND LI, D. From eliza to xiaoice: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering* 19, 1 (2018), 10–26.
43. WICAKSONO, A., HENDLEY, R. J., AND BEALE, R. Using reinforced implementation intentions to support habit formation. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI EA '19, ACM, pp. LBW2518:1–LBW2518:6.
44. WIRATUNGA, N., WIJEKOON, A., PALIHAWADANA, C., COOPER, K., AND MENDHAM, V. Fitchat: Conversational ai for active ageing.

45. XU, Y., AND WARSCHAUER, M. Young children’s reading and learning with conversational agents. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–8.
46. YANG, Z., DAI, Z., YANG, Y., CARBONELL, J., SALAKHUTDINOV, R. R., AND LE, Q. V. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems* (2019), pp. 5753–5763.
47. YU, Q., NGUYEN, T., PRAKKAMAKUL, S., AND SALEHI, N. “I almost fell in love with a machine”: Speaking with Computers Affects Self-disclosure. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–6.
48. ZHOU, M., QIN, Z., LIN, X., HU, S., WANG, Q., AND REN, K. Hidden voice commands: Attacks and defenses on the vcs of autonomous driving cars. *IEEE Wireless Communications* 26, 5 (2019), 128–133.
49. ZIMMERMAN, J., FORLIZZI, J., AND EVENSON, S. Research through design as a method for interaction design research in hci. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2007), pp. 493–502.