Designing for Trust

Exploring Trust and Collaboration in Conversational Agents for E-commerce

Meriç Dağlı

Designing for Trust

Exploring Trust and Collaboration in Conversational Agents for E-commerce

Meriç Dağlı, Author

Dan Lockton, Primary Advisor

Daragh Byrne, Secondary Advisor

A thesis submitted to the School of Design, Carnegie Mellon University, for the degree of Master of Design in Interaction Design

To My Beloved Family

Sevgili Aileme

Thanks

Istanbul Exporters' Association and Republic of Turkey, Ministry of Economy for providing a generous scholarship, which supported my education in Carnegie Mellon University, School of Design.

kynamatrix Research Network for supporting this research through their "Innovation through Collaboration" 2017 Grant Award.

Dan Lockton and Daragh Byrne for their invaluable guidance.

For their conversations and support, which made this possible:

Adem Önalan
Ahmed Ansari
Ahmet Burak Aktaş
Austin Beer
Austin Seungmin Lee
Ben Ginger
Bori Lee
Chirag Murthy
Chris Arrowood
Çiğdem Erbuğ
Deniz Camus
Devika Khowala
Dixon Lo
Eren Bali
Gautam Bose
Gordon Tindall
Gökçe Altun
Göktuğ Duman
Gülay Hasdoğan
Gülşen Töre Yargın
Güzin Şen
Jesse Wilson
John Zimmerman
Kimberly Harvey
Koray Benli
Ling 'Leah' Jiang
Mackenzie Cherban
Manjari Sahu
Manya Krishnaswamy
Matt Prindible
Merthan Öztürk
Molly Wright Steenson
Monique Smith
Nurie Agnes Jeong
Pablo Stanley
Rıza Selçuk Saydam
Sarah Marie Foley
Serdar Ariçelik
Shruti Aditya Chowdhury
Qian Yang
Vanessa Calderon
Vikas Yadav

Abstract

Trust can be thought of as one of the currencies that humans use to accept a technology in their everyday lives. One of those technologies are conversational agents (also known as chatbots) like digital assistants in our smartphones. To provide a more personal experience, increasing number of businesses are developing virtual agents with conversational interfaces by personifying their products and services through automation. While some agents are generalists like voice-based personal assistants, many of them are specialist agents that are designed for specific tasks in different domains including e-commerce. As user expectations get more complex each day, a collaboration between specialist agents is needed. In such a scenario, users' trust level may change dynamically due to the agent hand-offs. Additionally, despite many offered benefits, many people hesitate to trust conversational agents with complex tasks in certain e-commerce scenarios such as travel booking.

This design research project explores trust issues with virtual e-commerce agents in several collaboration scenarios between generalist and specialist agents and provides design guidelines for more trustworthy conversational agents.

Contents

Conversational Trust Design Checklist

Be Transparent

This thesis project explored trust and collaboration in conversational agents with the goal of identifying interaction design patterns that foster trust for building better relationships with users.

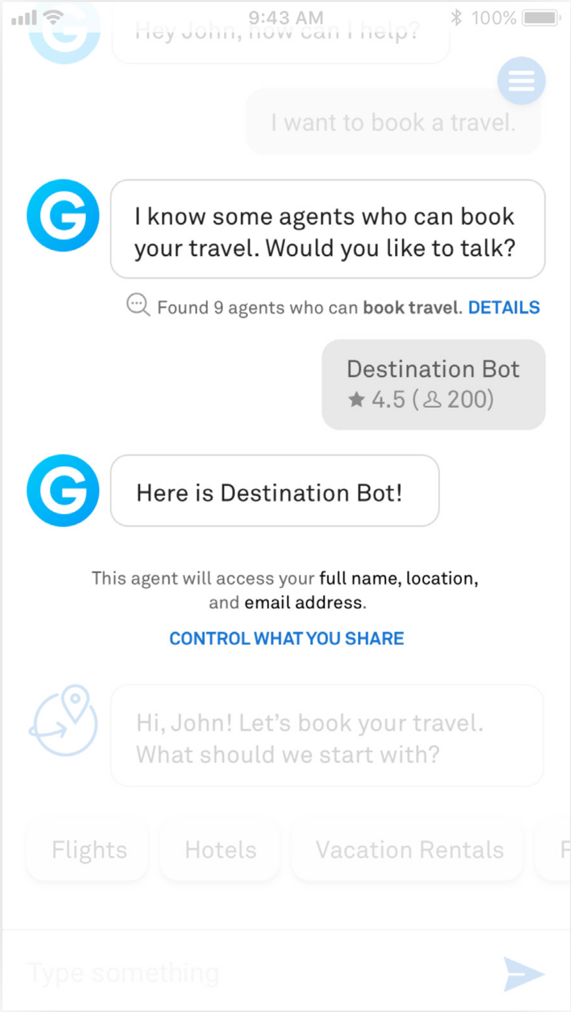
To explore what a collaborative future may hold for agents, it compared two scenarios with a travel booking conversational agent system through a Wizard of Oz prototype:

A negotiation scenario, which a meta agent did bargain on behalf user with other agents.

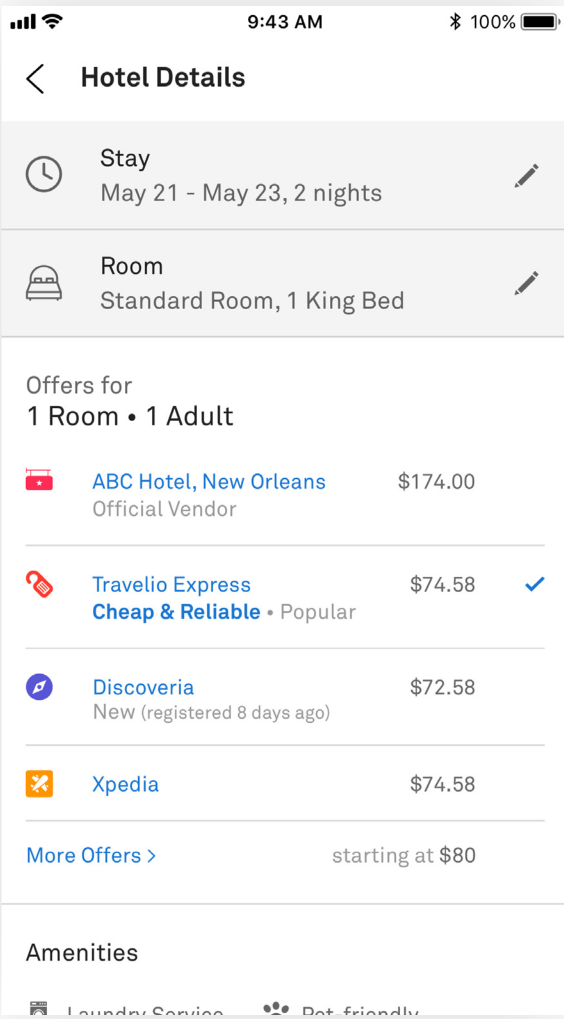
A bot-to-service composition where users interacted multiple agents for specific tasks.

Findings are synthesized into implications for interaction designers in five main themes.

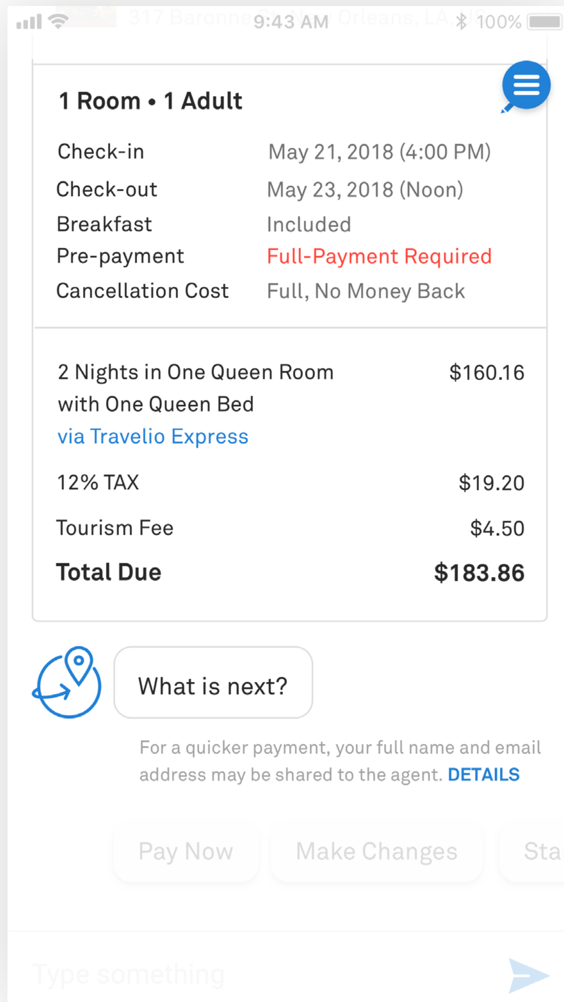
☐ Share What Agents (Need to) Know About User



☐ **Refer Others Cautiously, Visualize Confidence Level**

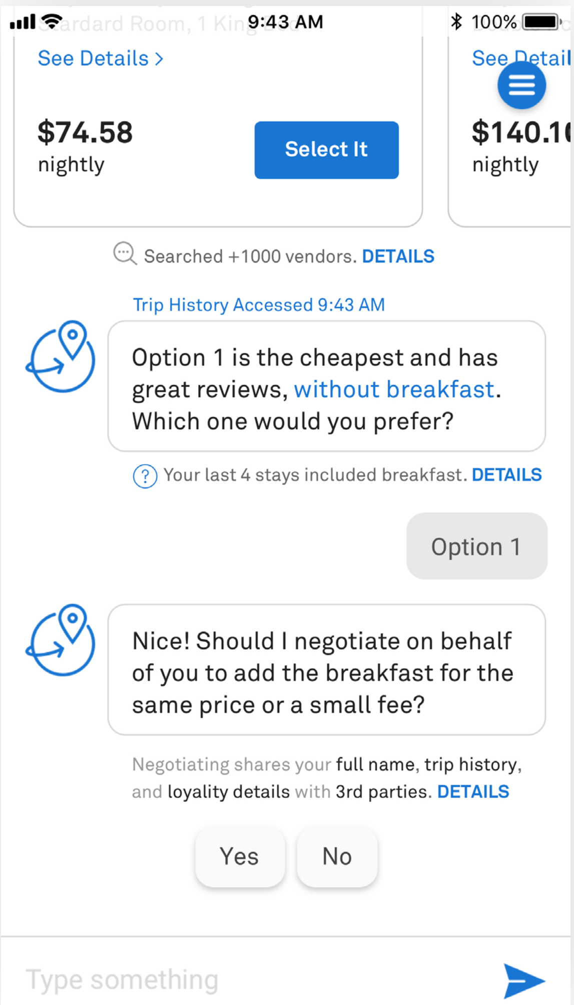


☐ **Give Specific Feedback to Clarify**

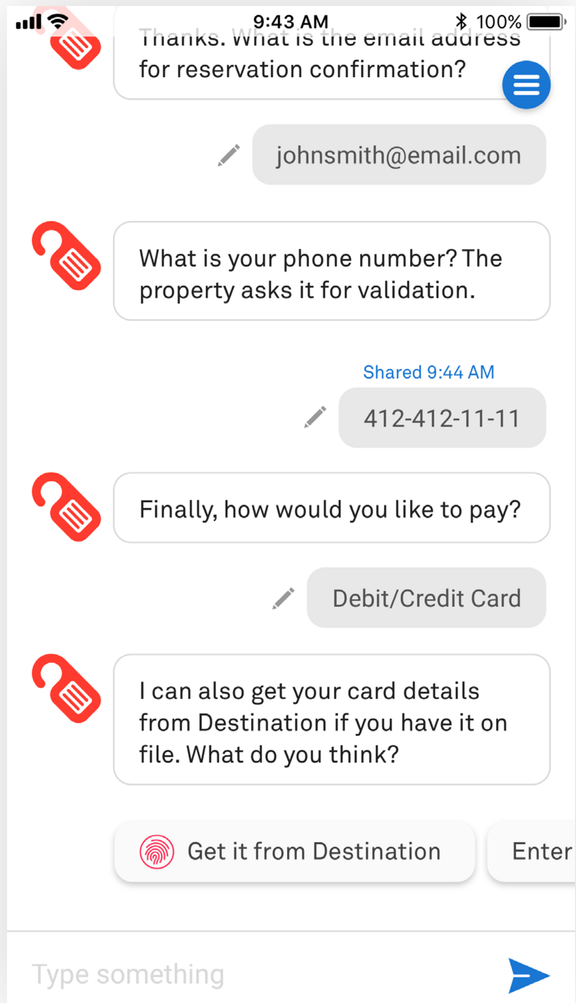


Give Control to the User

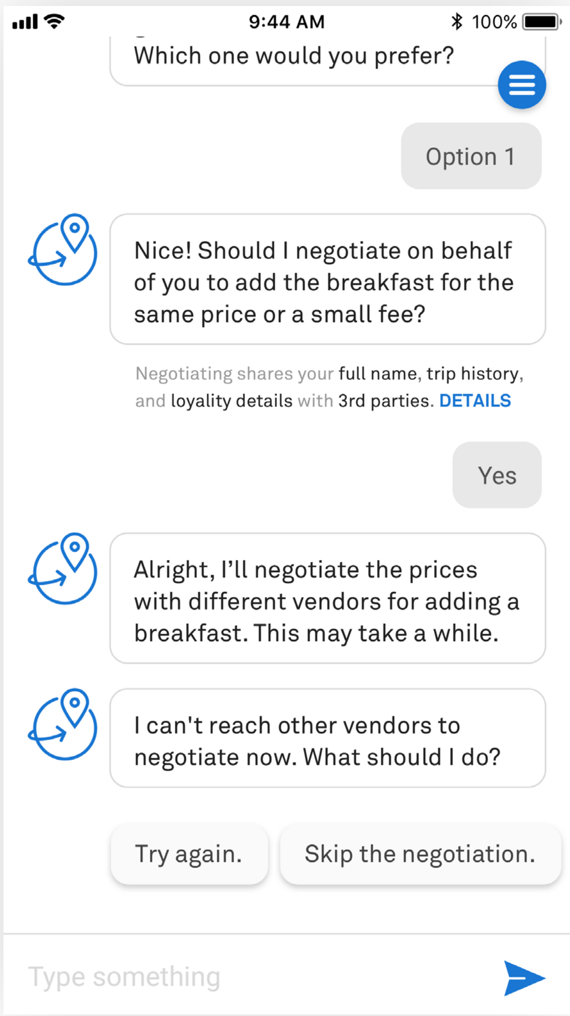
☐ Enable Users to Review Agent's Decision-Making



☐ Provide a Room for Revisions



☐ Fail Gracefully, Offer Auto-Recovery

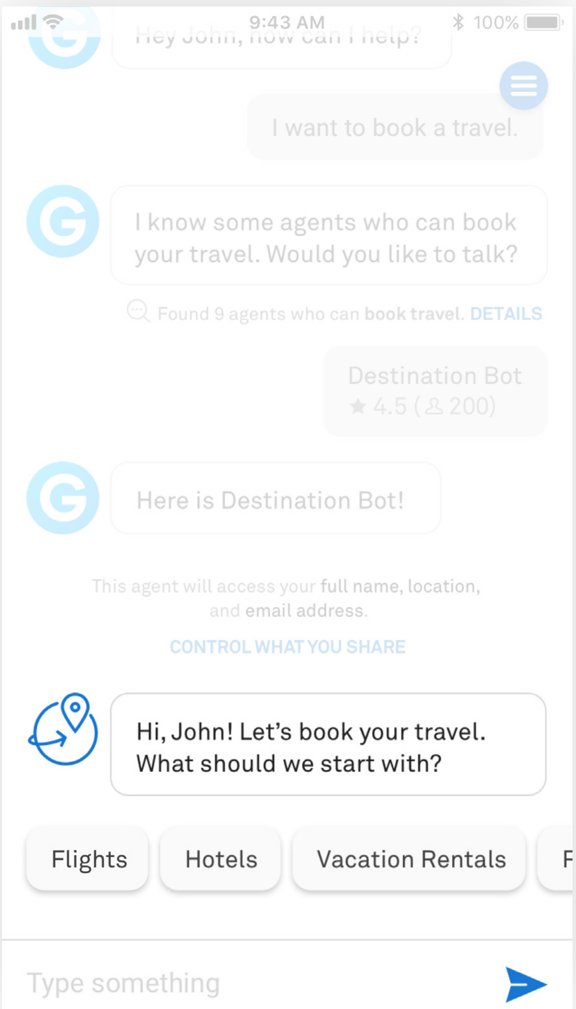


☐ Provide Alternatives for Agents

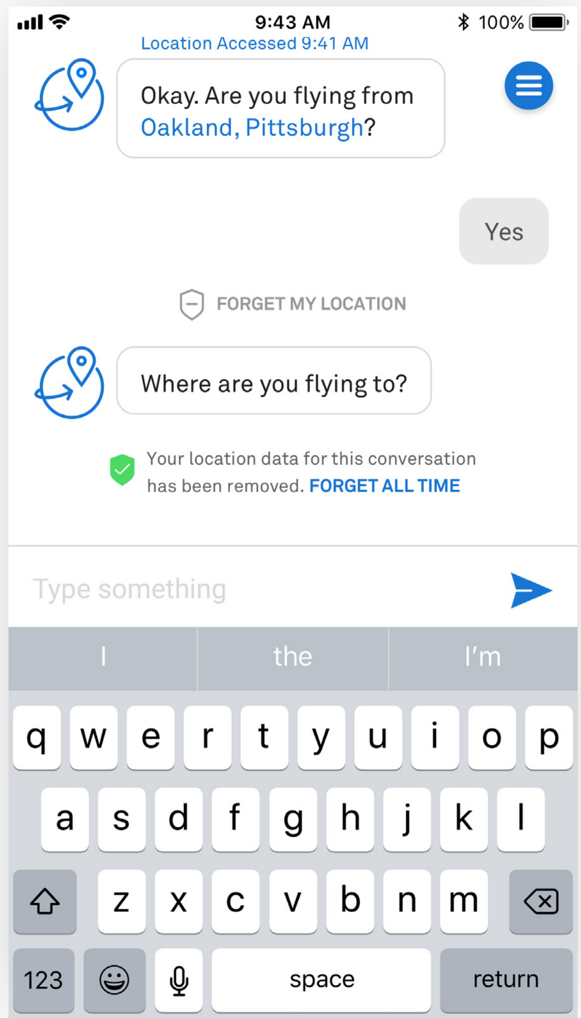
Non-Conversational Interfaces

Be Relevant

☐ Set the Expectations

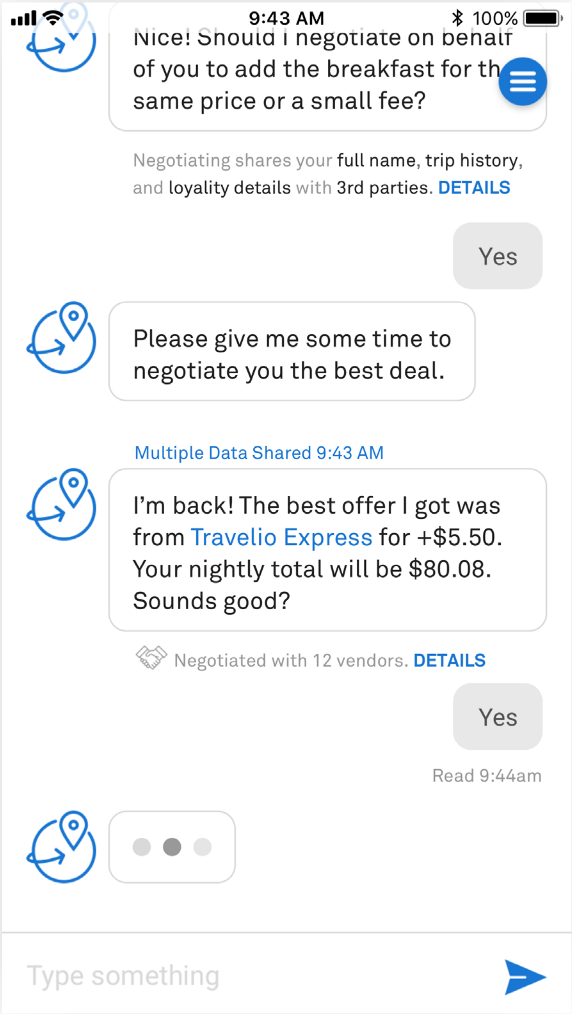


☐ Remember the Context and Forget it When Asked

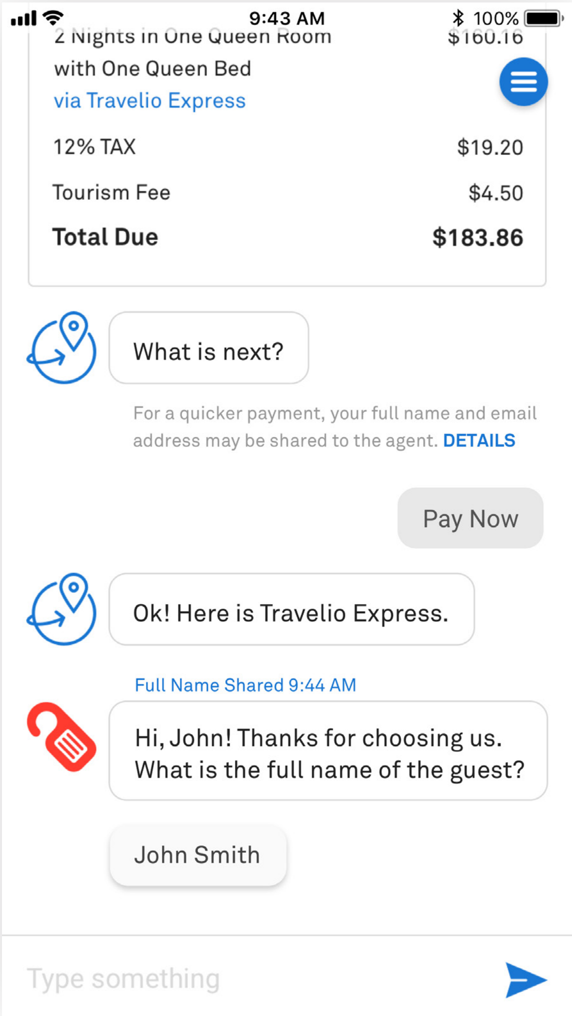


Be Responsive

- ☐ **Indicate the Writing and Processing Visually**

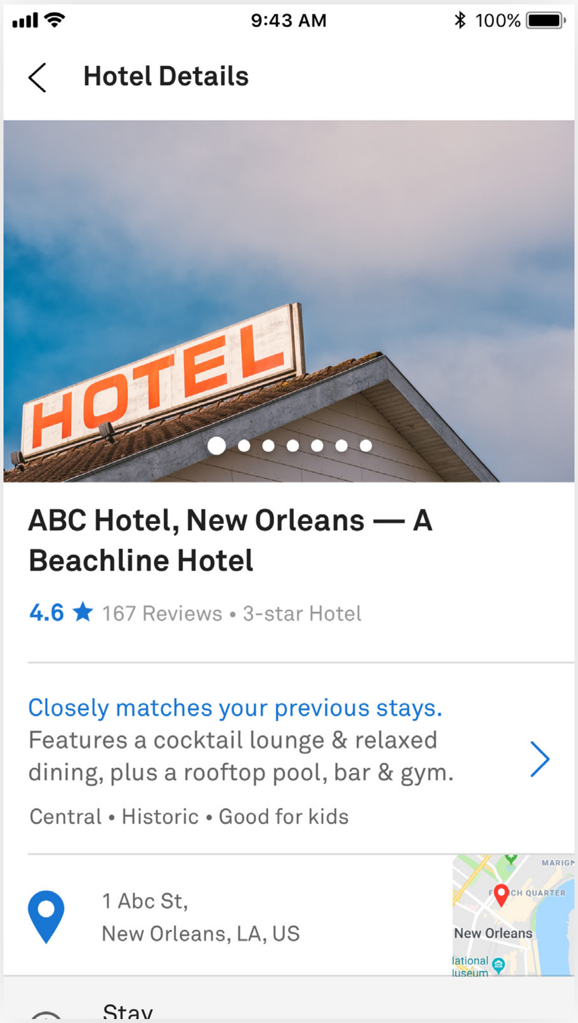


- ☐ **Do Not Indicate Hand-offs**

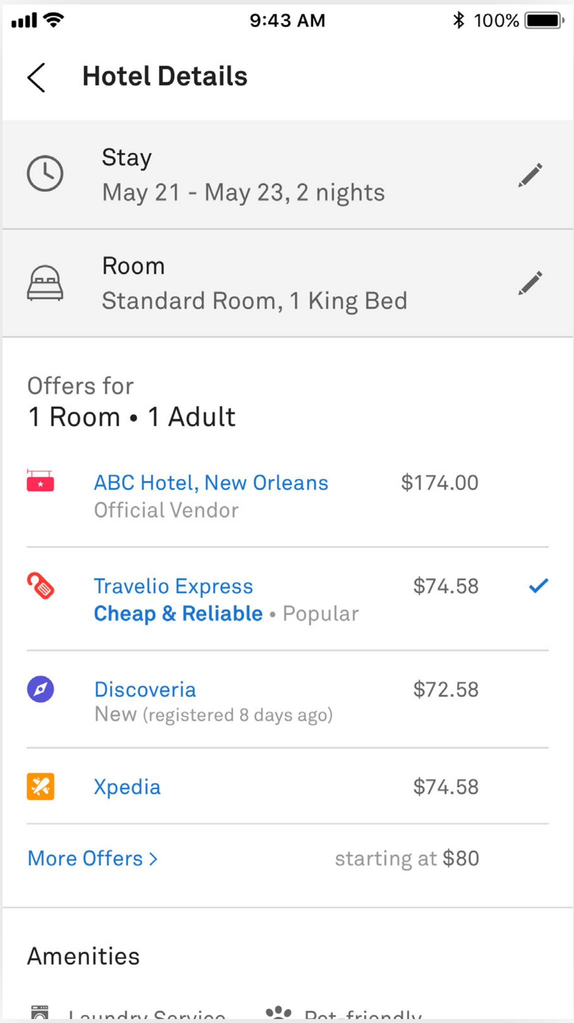


Be Visual

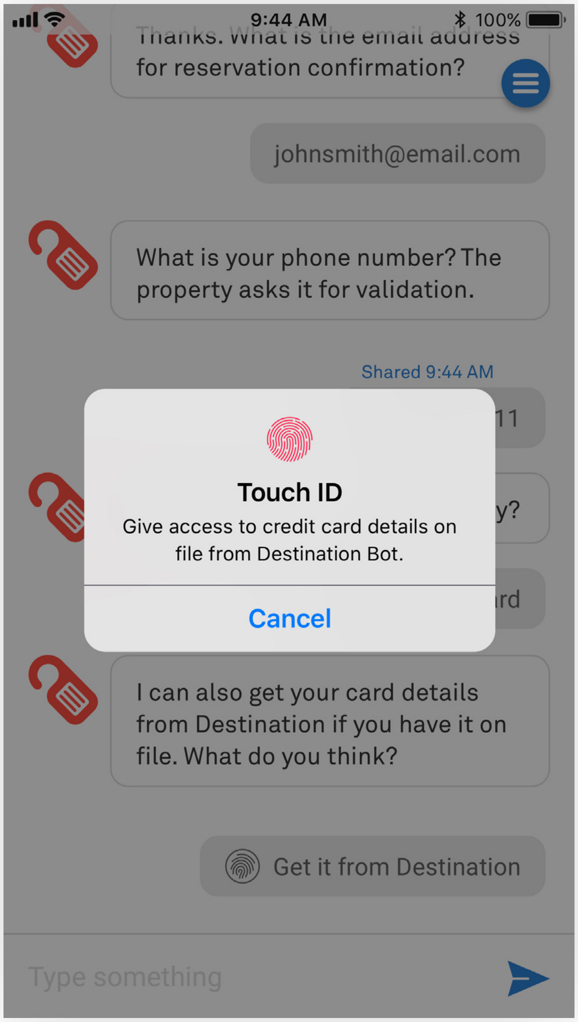
- ☐ **Use Visual Elements to Increase the Credibility**



- ☐ **Include Branding Where Possible**



☐ Provide Secure Gateways



The following chapters document the design process that lead to these implications and give more detail. As trust on conversational agents becomes more crucial each day with the advancement of computing, this project aims to continue exploring trustworthy design patterns with the feedback of whom will use it: designers.

Designing interfaces with using this checklist can help designers to build more trustworthy text-based conversational agents.

If you are a designer who is interested in collaboration, or wants to give feedback, please join the conversation.

Meriç Dağlı
mericdagli.com/designing-for-trust

Introduction

Conversational Agents: Focusing on Text-Based Agents

Each day our computers are becoming more like us: more humane and social. Since the 1960s, we have been communicating with more and more computer (programs) in our language. These programs that we interact with by talking or writing are called *conversational agents*, also known as dialog systems. It is estimated that more than 600 million people used voice-based conversational agents such as Apple Siri (2011), Google Assistant (2016), Amazon Alexa (2014), and Microsoft Cortana (2014) at least once every week in 2017[1]. As humans can converse different ways, using sounds, text, and gestures, agents are also designing to be multimodal. While they can be text-based, voice-activated or even embodied, the scope of this thesis is text-based conversational agents—also known as chatbots.

Today: Conversational Agents as Assistants

Besides 'chit-chatting,' text-based conversational agents help us with their domain of expertise as specified by their developers. For example, we have personal financial assistant chatbots that aim to help us with tracking our expenses[A] , mental health assistants that monitor our psychological health[B] , or even legal advisor chatbots that automatically appeal our speed tickets[C] . Alternatively, some agents try to become generalists by acquiring multi-domain skills. These agents set our alarms, remind us to take our clothes, do web searches for us, or suggest restaurants and so on. By 2022, we may have over 870 million devices that run conversational agents, which is expected to significantly increase end-user exposure to conversational agents in everyday scenarios.[2]

---

[1] Cherian, Elizabeth, and Jeremy Pounder. "Speak Easy Global: The Futures Answers to You," 2017.

[2] Botsman, Rachel. *Who Can You Trust? : How Technology Brought Us Together : And Why It Could Drive Us Apart*, 2017.

Future: Collaborative Conversational Assistants

As computing becomes social, conversational agents may accomplish more complex tasks by collaborating among themselves.[3] While today's agents mostly don't know what other agents are 'capable' of, tomorrow's agents will use knowledge and skill databases, which enable them to know what other agents can do. If an agent does not have enough expertise, it may refer other agents to help the user and sustain the conversation. In other words, whether the hand-off between agents is visible or not, we are likely to meet or be introduced to different agents along the way. In such a future, we may get to know other agents through the agents that we have already forged a relationship.[4] The state of our relationships with both agents, will depend on one core human trait: trust.

Trust in Conversational Agents

"Trust enables us to feel confident enough to take the risk and form a relationship with another entity while opening ourselves to be vulnerable."[5] In our relationships with computer programs, trusting them makes our complex experiences simpler.[6] When we trust a virtual agent, we believe in their caring intentions towards us and more recently perhaps we also have to believe in their decision-making skills. Naturally, we trust an agent to provide the outcome for which it initially set our expectations. For example, if a conversational agent advertised to us that it could set our alarm,

---

[3] Shevat, Amir. "5 Scenarios for How Humans and Bots Work Together." *Slack Platform Blog*, 2016. https:/medium.com/slack-developer-blog/5-scenarios-for-how-humans-and-bots-work-together-a0e636053ef4.

[4] Aboluwarin, Pelumi. "Chatbots — Igniting Division of Labour in AI." *Medium*, 2016. https:/medium.com/@pelumi/chatbots-igniting-division-of-labour-in-ai-1430fcc85c8d.

[5] Botsman, Rachel. *Who Can You Trust? : How Technology Brought Us Together : And Why It Could Drive Us Apart*, 2017.

[6] Ibid.

when we ask for it to “set the alarm” we have no reason to believe it will not set the alarm. Formed by our similar experiences with interfaces, we expect to receive a confirmation about the time of the alarm, and finally when the designated alarm time arrives, a ring from the agent to notify us.

It is important to note that, our interaction with agents is most of the time personal and private. Delegating our agency to them also means that we trust them that they can keep secrets, responsibly use and store our data and information. While our data has been shared more than ever, the collaboration between agents may also raise questions: *How will our data being disclose to the agents that we do not know and trust yet? Will our trust with one agent transfer to another one?* The answers to these questions depend on the context in which we are trusting the agent.

Where does trust in conversational agents matter most?

Trust helps humans to take action despite the risk and uncertainty. Although conversational agents’ black-box nature already possesses a risk and uncertainty for humans, our trust in them matters more in high stakes contexts, where there are higher risk or uncertainty. We want to make sure that we will not lose or undermine our health, our money, our family, our future, and so on. This thesis takes e-commerce as an example domain to demonstrate design experiments as it is a ‘high stakes’ domain, which many conversational have been built, but may not have been successful due to trust issues. While money usually becomes the highest stake in an e-commerce scenario, its transactional nature makes it a good case environment for the scope of this project: trust and collaboration in conversational agents.

Why travel?

As e-commerce is a broad category, this project also scopes down to a deeper level of e-commerce. Being a fragmented experience that requires different actors to collaborate, the travel booking journey has been used as an example case-study to explore how trust dynamics change in a collaboration scenario.

Scope

Starting from the first decision to converse, to relying on it repeatedly to complete tasks and activities, humans use trust to feel confident about their expectations towards agents.[7]

This project started by exploring trust in conversational agents as if it is a binary concept, implying trust is either present or not, with no grey areas. As the project developed, it acknowledged the contextual nature of trust. When users trust a conversational agent, in reality, they trust it in specific tasks and domains. Today, users trust 'skillful' agents who can help them with multiple and smaller tasks such as setting the alarm or calling a cab. If these agents do not understand users or if users' requests were out of their scope, users tend to lose their trust on these agents for these specific tasks as they do with humans.[8] [9] [10] One of the approaches to solve this issue is establishing a way of collaboration between agents.

In an agent-to-agent collaboration scenario, an agent refers users to another agent with different expertise and knowledge from its trusted network when it cannot help users by itself. Whether users experience these hand-off moments during these agent referrals or

[7] "Trust and Power, Niklas Luhmann, Wiley (1979)"

[8] Botsman, Rachel. *Who Can You Trust? : How Technology Brought Us Together : And Why It Could Drive Us Apart*, 2017.

[9] Bhattacharya, Rajeev, Timothy M. Devinney, and Madan M. Pillutla. "A Formal Model of Trust Based on Outcomes." *The Academy of Management Review* 23, no. 3 (July 1998): 459.

[10] Coleman, James S. *Foundations of Social Theory*. Belknap Press of Harvard University Press, 1990.

not, the collaboration between agents raises many questions about trust and its transitivity from the agent they already know to the other agent. To find opportunities for designing more trustworthy chatbots, the focus of this project is how the collaboration between multiple conversational agents affects the users' trust.

The Goal and Research Questions

This thesis aims to give actionable suggestions for practioners to form or build users' trust through design.

The success of this thesis depends on its ability to explore and identify opportunties in the literature with a designers mindset.

To achieve this aim and drive success, it explores the following research questions:

- How might we design to foster users' trust in conversational agents towards a future where agents collaborate with each other?
- How might we design to foster users' trust in conversational agents?
- How will agents collaborate?
- How does a collaborative agent scenario affect the user experience of conversing with a conversational agent?
- How will collaboration influence users' trust?
- In a collaboration would users trust transfer between agents that they already trust, and they do not know?
- How does the behavior of stranger conversational agents' affect the overall perceived trust level of the user experience?
- What happens if the stranger conversational agent gives inconsistent information to users? Does this affect users' trust in the initial agent?

Methodology

The nature and the complexity of the research themes need insights from different channels. Therefore this project combines desk research, user research, and research through design methodologies to tackle the research questions from different aspects. It reviews current literature and identifies related works within the scope of the project to identify gaps, then uses qualitative user research methods such as surveys, interviews, generative workshops, and artifact evaluations to pull insights. Then it synthesizes the iterated learnings into a conversational trust design checklist and best practices with the aim of providing actionable design suggestions for practitioners.

Background

Literature Review

This chapter reviews related literature on trust, conversational agents, and collaboration. This review aimed to understand and frame trust, conversational agents, and trust in conversational agents to identify opportunity spaces for exploratory research and useful directions for final design implications.

This review starts with tracing back what makes a conversational agent trustworthy, by defining trust and the conversational agents. Then, it continues with exploring current challenges of trust and task-based conversational agents with an outlook to future opportunities by reviewing the collaboration between agents. To move further, it classifies agents based on their task capabilities and area-of-expertise: generalist agents (also known as meta-agents) and specialist agents (also known as experts agents).

After modeling the collaboration in generalist and specialist agents, this review continues with examining the literature on agent collaboration scenarios and its relationship with trust.

This review concludes with areas for reviewing related works and opportunities for exploratory research.

Conversational Agents

As a society, we have labeled a human-like virtual character in many different ways.[11] While we label computers that we talk, text, interact interchangeably as chatbots, virtual agents, virtual assistants, virtual companions, avatars, 'artificial intelligence' or conversational agents, these labels can also be used to define subtle differences.[12] For example, text-based conversational agents often are called chatbots. Throughout this thesis, the terms above will refer to the same definition of conversational agent: a computer program that interacts with humans in natural language.[13]

Fig 1. A conversation with ELIZA (from: https:/github.com/maldiny/Chatbots-en-Castellano)

The realm of conversational agents started with the first known chatbot, Eliza, which was a computer program that imitated a Rogerian psychotherapist with a text-based interface in 1964.[14] Eliza's creator, Joseph Weizenbaum was able to create one of the first programs that managed to imitate and create an illusion of

[11] Luger, Ewa, and Abigail Sellen. "'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents." *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2016, 5286–97.

[12] Yorick Wilks. 2010. Is a companion a distinctive kind of relationship with a machine? .In Proceedings of the 2010 Workshop on Companionable Dialogue Systems (CDS '10).

[13] Bayan Abu Shawar and Eric Atwell. 2007. Chatbots: are they really useful? In LDV Forum, 29–49.

[14] Weizenbaum, Joseph. "ELIZA — A Computer Program For the Study of Natural Language Communication Between Man And Machine." *Communications of the ACM* 9, no. 1 (1966): 36–45.

the human-human interaction in human-computer interaction. 15 years before the personal computer became mainstream, he was surprised to find out how powerful the computer's programs, or in other words algorithms, become when they interact with humans in their language. Nearly, all people who tried Eliza thought that it was intelligent enough to truly understand them, while Weizenbaum was using an early natural processing technique with recognizing minimal context from users' response.

Almost 55 years later from Eliza, today, as humans we continue to think that computers behave like humans. Advances in areas such as machine learning and natural language understanding continue to redefine our relationship with computer programs as more human-like social actors in our lives. Whether we use them for our primary-care appointments (Carbon Health - Carbon Bot[15]), to wake up us at a specific time (any voice-based assistant) or even to call restaurants to make reservations (Google Assistant), we see conversational agents as digital assistants to support us in completing tasks. They are learning more about us to act on behalf us in both virtual and real worlds.

As the capabilities of conversational agents increase day-by-day, one can classify agents based on their developed area-of-expertise: generalist agents and specialist agents.

Mostly developed by larger technology companies such as Apple, Google, Amazon, and Microsoft, generalist agents help users with accomplishing multiple tasks and route them to others. Respectively, Siri, Assistant, Alexa, and Cortana are the generalist agents of the companies above. Being able to understand more user intents, these agents, generally, aim to become personal assistants of the users. Users can ask them to do relatively generic tasks such as setting the alarm, doing a web search, playing music, calling a cab, scheduling meeting. In the literature, generalist agents are also

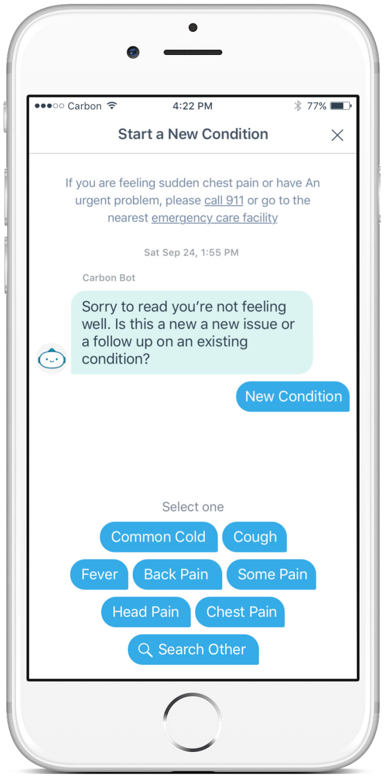


Fig 2. CarbonBot from Carbon Health, onboards a patient.
(from https:/carbonhealth.com/press)

[15] Carbon Healthcare. "A Fresh Take on Health Care," 2017. https:/carbonhealth.com/ehr.

referred as super bots[16] or meta-agents[17], as they often combine multiple tasks and the functionality of specialist agents in one agent. So what is a specialist agent?

Specialist agents have a particular area of expertise. They are capable of handling highly focused and particular tasks such as DoNotPay bot, which can give free legal advice[D], Icelandair's bot which can book flight tickets[E], or Digit, which saves money automatically by setting money aside from users bank account every other day based on their goals[F]. They can even help emergency call line representatives to detect if a person has a cardiac arrest in an emergency, like Corti AI.[G] As diverse as their functionality can be, these agents also use a wide range of platforms and interfaces. While some of them such as DoNotPay Bot, Digit, or Corti have a dedicated user interface, most of the specialist agents are parts of the ecosystems around generalist agents or bot platforms such as Actions on Google[H], Amazon Alexa Skills[I], Microsoft Cortana Skills[J], Apple SiriKit[K], and Facebook Messenger[L]. For example, another specialist agent, 1-800-Flowers Assistant, which helps users to send flower arrangements, is accessible through multiple platforms such as "Actions on Google" and "Alexa Skills".

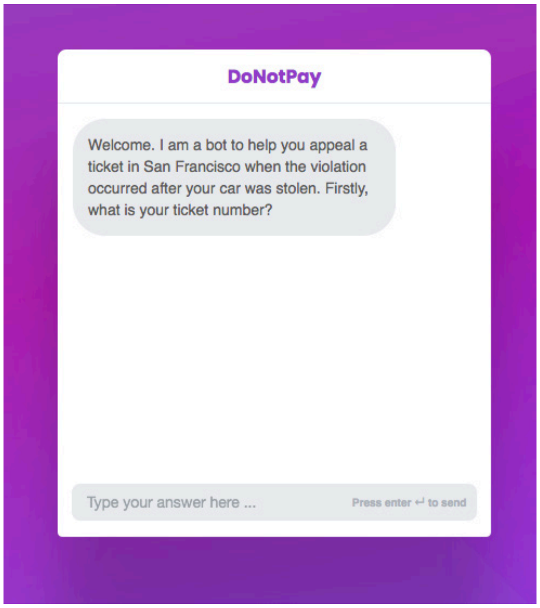


Fig 3. A conversation with DoNotPay bot (from: https:/www.donotpay.com/)

Living inside the platforms of generalist agents also increases the traction and the engagement of the specialist agents by increasing agent discoverability through referrals and collaboration in two distinct approaches. In the first approach, users have to explicitly ask the generalist agent by the specialist agent's name to converse with it. These call-by-name intents are called explicit invocations.[18] For example, in Google Assistant ecosystem, 1-800-Flowers becomes

[16] Amir Shevat. "Designing Bots. Creating Conversational Experiences," 2017.

[17] Aboluwarin, Pelumi. "Chatbots — Igniting Division of Labour in AI." *Medium*, 2016. https:/medium.com/@pelumi/chatbots-igniting-division-of-labour-in-ai-1430fcc85c8d.

[18] Google. "Explicit Invocation." *Actions on Google*, 2017. https:/developers.google.com/actions/discovery/explicit.

an “Action on Google,” which users access it by saying “Okay, Google talk to 1-800-Flowers.” from Google Assistant. Similarly, users can also access 1-800-Flowers agent by asking it with its name from Alexa, Amazon’s generalist agent. This way, 1-800-Flowers Assistant, a specialist agent, becomes a skill for users of generalist agents, which they can reach without leaving their interface of the generalist agent, yet users have to ask it explicitly. On the other hand, while it is an emerging idea, agents may also refer specialist agents if they know they can handle users’ request. These are called implicit invocations.[19] For example, when a user says “I want to send flowers to my mother” Google Assistant may refer and invite ‘1-800-Flowers Assistant to the conversation. In contrast, rather than inviting a third party agent to the conversation, a generalist agent can also communicate with third party agent on the back-end, without surfacing the actual communication to the user as the way Apple Siri does it currently. For example, when users ask to call for a cab, Siri conveys the information from the user to the apps of the ridesharing platforms such as Uber/ Lyft on the backend. Then it conveys the response back to the user without inviting the third party to the conversation. An example reply from Siri would be “Uber told me that it can get you an Uber in 10 minutes.”.

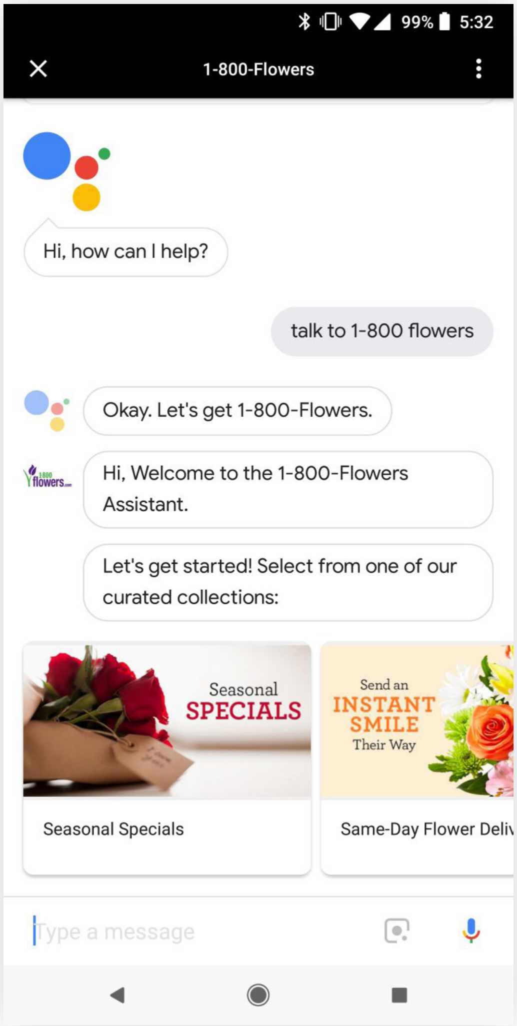


Fig 4. Google Assistant refers 1-800-Flower assistant through an explicit invocation. (from: https:/9to5google.com/2018/02/07/how-to-order-1-800-flowers-using-google-asstiant-android-basics/)

While agent referrals help generalist agents to be even better, they can also enable the discoverability of other agents.[20] For example, when users buy tickets for an outdoor show, a ticket agent can collaborate with a Weather agent to see if users would need a coat or an umbrella during their event.

As we see referrals between agents are emerging, there are not many real-world examples of *continous referrals*, which will form a referral-chain between more than two agents. Longer referral chains may not be preferred due to lack of existing

[19] Google. “Implicit Invocation.” *Actions on Google*, 2018. https:/developers.google.com/actions/discovery/implicit.

[20] Amir Shevat. “Designing Bots. Creating Conversational Experiences,” 2017.

communication protocols. The related works chapter includes some examples of these multiple agent collaborations.

Whether they surface handoffs or not , conversational agents can collaborate in many ways. Two of these ways are either through searching what type of user intents other agents are capable of fulfilling in ‘an intent database’ or simply asking another agent that works truly as a meta agent, an agent that knows all other agents. Besides agent-centered solutions, there are also ecosystem solutions such as Botchain, which is a decentralized platform for bot transparency, compliance and collaboration.[21] While today’s generalist agents are slowly becoming meta-agents and decentralized solutions are being developed, knowing what other agents are capable of, and their intents, are hard problems to tackle.

The success of the agents depends on one human mental state that can’t be automated: trust.[22] An important outcome lies in the heart of the conversations of agents of today and future: building trust.[23] “If we do not trust these conversational agents at the first place, there is no point building them.”[24] says Rachel Botsman, the author of *Who can you trust?* (2017).

She also argues that “it is why developers and designers are designing interfaces, based on metaphors to earn users’ trust in the first place, including manipulating the appearance.”[25] While this argument positions trust in “robots” with design, understanding what a conversation is, and what trust is, are essential before discussing trust in the agents and its possible collaborative future.

---

[21] Botchain. “Can You Trust AI?,” 2018. https:/botchain.talla.com/.

[22] Botsman, Rachel. *Who Can You Trust? : How Technology Brought Us Together : And Why It Could Drive Us Apart*, 2017.

[23] Ibid.

[24] Ibid.

[25] Ibid.

Conversation

What is conversation?

Conversation is more than just an exchange of information; it is how we work and live together for more than 100,000 years.[26] It is the mutual progress of negotiation and collaboration to create and agree upon meanings to first come to a mutual understanding and then operate together. While it is difficult, we have to converse to advance, to learn new concepts, share and evolve the knowledge, and come to an agreement.[27]

To explain how conversation works, specifically about interaction design, Dubberly and Pangaro[28] break it into six small rules:

1. *"**Open a Channel:** To set up a common-ground, Participant A takes the leap and sends a message to the Participant B.*
2. ***Commit to Engage:** Participant B show interest to join.*
3. ***Construct Meaning:** Connection between participants start to build via commonalities and ideas. As conversation builds up, participants start to assign meaning to the interaction.*
4. ***Evolve:** Participants either or both learned or gain something such as new beliefs, relationships, decisions, or ideas.*
5. ***Converge on Agreement:** Participants discuss their understanding until they are aligned. If things go well, they may reach an agreement or they have to resolve the situation.*
6. ***Act or transact:** An exchange happens between participants. Then, they take a tangible action or achieve a mental goal."*[29]

A conversation nevertheless has many limitations including the infrastructure within which it happens, the limits of participants

[26] Glaser, Judith E. *Conversational Intelligence : How Great Leaders Build Trust and Get Extraordinary Results*, n.d.

[27] Beer, Austin. "Let's Chat!" Hyper Island, 2016.

[28] Dubberly, Hugh, and P Pangaro. "What Is Conversation? How Can We Design for Effective Conversation?" *Interactions Magazine* XVI, no. August (2009): 1–9.

[29] Ibid.

who are pursuing it or how much they cooperate.[30] For example, in a loud environment, the conversation may not go as expected due to the noise or with specific participant groups that have limited capacity, a conversation may go nowhere. A conversation can also break down if participants do not cooperate enough. Then, *how much co-operation does enough for a conversation?*

What makes a conversation better than others?

Not all conversations are equal; some are better than others. After studying what makes conversations more effective, the linguist Paul Grice proposed *the cooperative principle* along with four maxims, called *Gricean maxims*. He argues that people will join a conversation with a shared goal or purpose.[31] Based on the assumption that there is a cooperation between participants, the following maxims describe traits of effective communication:[32]

*"**Maxim of Quality:** Being co-operative on the truth of information that is used to converse. Not conversing what believed to be false. Not conversing things without adequate evidence.*

***Maxim of Quantity:** Being co-operative on the quantity of information that is used to converse. Giving information as much as required to pursue the conversation. Not providing excessive details, making the conversation more informative than needed.*

***Maxim of Relevance:** Being co-operative on the relevance of the information that is used to converse. Giving appropriate and closely-connected information to the conversation.*

***Maxim of Manner:** Being co-operative on how we attempt to converse. Communicating clearly without any obscurity and ambiguity with following the order conversation needs."*[33]

[30] Dubberly, Hugh, and P Pangaro. "What Is Conversation? How Can We Design for Effective Conversation?" *Interactions Magazine* XVI, no. August (2009): 1–9.

[31] Grice, Paul. "Logic and Conversation." *Syntax and Semantics*, 1975.

[32] Ibid.

[33] Ibid.

Gricean Maxims provide rules for effective conversations between humans, yet humans can fail to observe a maxim in many ways.[34] Often these failures can generate more complex meanings.[35] For example, flouting maxims is a widely used way to create humor in comedy. Sheldon, a character in the TV series, *The Big Bang Theory*,[36] floats the maxim of quantity by giving unnecessary information on how ketchup is made when his friend asks him to pass the ketchup.[37] While this behavior adds to Sheldon's personality, it may have been found hilarious by the viewers.

Learning about how cooperation help humans to exchange ideas, coming to agreements, and take actions, sparked my curiosity. *How much of the human conversations do apply to agents?*

Conversations with Computers

Computers started to converse in the human language since Eliza while the goal of creating a machine that can exhibit behaviors indistinguishable from humans goes back to 1950s with Alan Turing's paper, "Computing Machinery and Intelligence."[38] A similar paradigm, "Computers are social actors" argues that "humans react to technologies that possess human-like behaviors or social cues, the same way they respond to humans."[39]

When a computer, or in this context, conversational agent shows human-like behaviors such as using a human language, turn-taking in a conversation and responding the same way a human would, users personify the technology. On the other hand, "when

[34] Grice, Paul. "Logic and Conversation." *Syntax and Semantics*, 1975.

[35] Juez, Laura Alba. "Verbal Irony and the Maxims of Grice's Cooperative Principle." *Revista Alicantina de Estudios Ingleses* 8, no. 8 (1995): 25–30.

[36] Beer, Austin. "Let's Chat!" Hyper Island, 2016.

[37] Millard, Miss. "Grice's Maxims in 'The Big Bang Theory.'" *Youtube*, 2014. https:/www.youtube.com/watch?v=vEM8gZCWQ2w.

[38] Turing, Alan M. "Computing Machinery and Intelligence." In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, 49:23–65, 2009.

[39] Lankton, Nancy, D. Harrison McKnight, and John Tripp. "Technology, Humanness, and Trust: Rethinking Trust in Technology." *Journal of the Association for Information Systems* 16, no. 10 (2015): 880–918.

people think that the other part is human than a computer, they tend to put more effort into their conversations and show more engagement." [40] "While humans expect politeness in human conversations, they dislike the excessive politeness and repetitions with a machine."[41] Overall, these mean that humans use their existing model of human-human conversation and follow the cooperative principle, social cues, and Gricean maxims when they 'converse' with conversational agents and expect the virtual agents and personas, to use them, too.[42] However, as the technology is working towards to the aim of becoming indistinguishable from humans on conversations with humans[43], using excessive anthropomorphism, designing technology that has human characteristics, raise ethical issues despite its usefulness. These technologies may be abused and may use with malicious intentions to deceive or discomfort humans. [44]

Fig 5. Agents are being design to have human-like qualities. Sophia, a humanoid robot from Hanson Robotics. (from: https:/www.flickr.com/photos/itupictures/35008372172)

By reviewing the literature on conversations and computers, I learned that conversational agents have to follow the same rules and conventions as humans, for interacting with humans. Before discussing what this means for the design of conversational agents, I also wanted to learn more about one fundamental question: why are we designing conversations for virtual agents?

My review of conversational design also pointed out a critical answer to these questions: to gain users' trust in technology.

[40] Luger, Ewa, and Abigail Sellen. "'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents." *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2016, 5286–97.

[41] Ibid.

[42] Giangola, James. *Applying Built-in Hacks of Conversation to Your Voice UI (Google I/O '17)*. Youtube, 2017. https:/www.youtube.com/watch?v=wuDP_eygsvs.

[43] Knight, Will. "Can This Man Make AI More Human? - MIT Technology Review." *MIT Technology Review*, 2015. https:/www.technologyreview.com/s/544606/can-this-man-make-ai-more-human/.

[44] Farokhmanesh, Megan. "Deepfakes Are Disappearing from Parts of the Web, but They're Not Going Away." *The Verge*, 2018. https:/www.theverge.com/2018/2/9/16986602/.

Trust

In the simplest definition, "trust is a confident relationship with the unknown."[45] It is a social construct that originates from interpersonal relationships. Therefore everyone has their definition, which makes it "the most defined sociological concept."[46] We have so many definitions also because trust is highly contextual.

One can also trust or distrust something generally, yet at the end, the outcomes in specific context and situations will inform trust. For example, people still buy newspapers while they may claim they cannot be trusted. In its core, people's trust in newspapers is formed by specific actions and behaviors of the newspapers in specific moments and contexts. Then, this informs their overall trust in newspapers as negative, although they still instinctively trust other parts of the newspaper so that they continue to buy it. This effect is called the trust paradox.[47]

All definitions of trust include a situation where there is an unknown outcome. A predictable unknown outcome introduces risk and a genuine unknown outcome introduces uncertainty.[48] Everyday life has full of unknown situations. For example, risk occurs when we are working with someone, who we know what they are capable of. Besides, uncertainty occurs when we work with a company with many different people and we cannot predict what all of the employees are capable. In both situations, our positive expectations of others push us to make ourselves vulnerable by expecting an outcome from them, and for them not to exploit us. If we swap the 'positive expectations' part with trust, the meaning would not change as trust is also defined as positive expectations towards another party's behaviors and intentions.

[45] Botsman, Rachel. *Who Can You Trust? : How Technology Brought Us Together : And Why It Could Drive Us Apart*, 2017.

[46] Ibid.

[47] White, Mathew P. "The Trust Paradox: The Role of Context Effects in Stated Trust Judgements." *Journal of Risk Research* 10, no. 7 (October 2007): 977–88.

[48] Knight, F. H. "Risk, Uncertainty and Profit, 1921." *Boston and New York* 23, no. 3 (1921): 135–57.

‘The Mechanics of Trust’ framework explains how trust is formed and built in three steps based on expectations:

1. *“Both parties receive trust-inducing signals from each other. Adding these signals to their beliefs and experience, trustor form a perception about trustee.*
2. *Based on its perception, trustor forms expectations with uncertainty and risk on trustee.*
3. *Trustee fulfills the trustor’s expectations.”*[49]

Fig 6. Dimensions of human-trust

Dimensions of human-human trust

Understanding how people ‘build trust’ as an additive process was important, yet it does not specify what signals and beliefs parties have to form their perception, in other words how we measure trust. Researchers often use these three ‘trust beliefs,’ to measure trust between humans:

“**Competence:** The belief that a person has the skills, competencies, and characteristics to influence in a specific domain.”[50]

“**Benevolence:** The belief that a person will want to do good another person, care him/her in addition to themselves.”[51]

“**Integrity:** The belief that a person adheres to set of principles such as honesty, promise-keeping.”[52]

While these beliefs hold the key to trust between humans, *are they also effective for measuring humans’ trust in a technology?*

---

[49] Riegelsberger, Jens, M. Angela Sasse, and John D. McCarthy. “The Mechanics of Trust: A Framework for Research and Design.” *International Journal of Human-Computer Studies* 62, no. 3 (March 1, 2005): 381–422.

[50] Mayer, Roger C, James H Davis, and F David Schoorman. “An Integrative Model of Organizational Trust.” *Source: The Academy of Management Review* 20, no. 3 (1995): 709–34.

[51] Mcknight, Dh, M Carter, Jb Thatcher, and Pf Clay. “Trust in a Specific Technology: An Investigation in Its Components and Measures.” *ACM Transactions on Management Information Systems* 2, no. 2 (2011): 1–12.

[52] Lankton, Nancy, D. Harrison McKnight, and John Tripp. “Technology, Humanness, and Trust: Rethinking Trust in Technology.” *Journal of the Association for Information Systems* 16, no. 10 (2015): 880–918.

Dimensions of trust in a specific technology

If one replaces a human with a specific technology in a trust relationship, three trust beliefs transform into these three beliefs:

"Functionality: The belief that technology has the capacity or capability to complete a required task. Similar to Competence

Helpfulness: The belief that technology will provide adequate help and guidance for a human to be successful excluding the moral agency and volition (i.e. will) that humans have. Similar to Benevolence

Reliability: The belief that technology will work consistently and predictably. Similar to Integrity"[53]

As trust is contextual, even this measurement, trust in specific technology, is not specific enough to conversational agents, which uses conversations to be more human. Therefore, trust in conversational agents falls in between these two set of beliefs.

Trust in Humans	**Trust in a Specific Technology**
Competence/Ability	Functionality
Benevolence	Helpfulness
Integrity	Reliability

Trust in conversational agents

While in its core a conversational agent is a computer program, the experience of using it feels like conversing with a human by having humanlike qualities. As conversational agents follow the same conversational path with humans, human-human trust mechanics also apply to them. First, users seek trust and trustworthiness cues from agents based on their experiences and beliefs. Second, they form an expectation of agents' capabilities with some level of risk and uncertainty. Finally, the agent fulfills users' expectation by providing an expected outcome.

[53] Mcknight, Dh, M Carter, Jb Thatcher, and Pf Clay. "Trust in a Specific Technology: An Investigation in Its Components and Measures." *ACM Transactions on Management Information Systems* 2, no. 2 (2011): 1–12.

Dimensions of trust in conversational agents

Since some trust is needed to try a conversational agent, gaining the initial trust of the user is important for conversational agents. If a conversational agent failed to provide trust signals initially, users might not accept its suggestions even if it revises itself later.[54] In contrast, if an agent was able to show trust signals in the first experience, people may continue to use it even if its reliability decreases.[55] Similarly, when users encounter an error in an algorithm such as conversational agent, they may stop using the system.[56] These show that signalling reliability (consistency and predictability) and trustworthiness is vital to gain users' initial trust (increasing their confidence level on the agent) as conversational agents tend to make recommendations and decisions. While trust cues are widely discussed in interaction design literature, one of them is specifically relevant to conversational agents: small talk.

Small talk is a conversation about things which are not important, often among people who do not know each other well. Small talk is a conversation that helps people to build rapport, credibility to achieve their interpersonal goals.[57] Similar to humans, small talk can also help conversational agents to gain the confidence of users towards them. While the contents of the small talk may not be contextually essential or relevant, they help humans to 'test waters' with someone they do not know. So humans also tend to test waters iteratively with a conversational agent to see if it is worth their trust and to build a relationship. They also tend to test boundaries of an agent's intelligence and level of understanding to influence their expectations by small talk.[58]

[54] Yuksel, Beste F, Penny Collisson, and Mary Czerwinski. "Brains or Beauty: How to Engender Trust in User-Agent Interactions." *ACM Transactions on Internet Technology (TOIT)* 17, no. 1 (2017): 1–20.

[55] Ibid.

[56] Dietvorst, Berkeley J, Joseph P Simmons, and Cade Massey. "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err." Accessed May 13, 2018.

[57] Beer, Austin. "Let's Chat!" Hyper Island, 2016.

[58] Medhi Thies, Indrani, Nandita Menon, Sneha Magapu, Manisha Subramony, and Jacki O'Neill. "How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians." In *Lecture Notes in Computer Science*, 10513 LNCS:441–59, 2017.

While increasing number of conversational agents are 'capable of' making small talk, a conversation with an artificial entity is still far from being technically perfect. The performance and behavior of a conversational agent are still prone to errors. Again, "there are no 'errors' in a conversation."[59] Users expect an agent to converse cooperatively and not convey what they cannot be capable of and yet, sometimes an error happens in an agent due to failing to match the user's response. Therefore, how errors are being handled in conversational agents affects users' trust.[60] While there also may be errors related to agents' intelligence level of understanding user intents, which directly affect the conversation, there are also behavioral errors, caused by an agent's intentions.

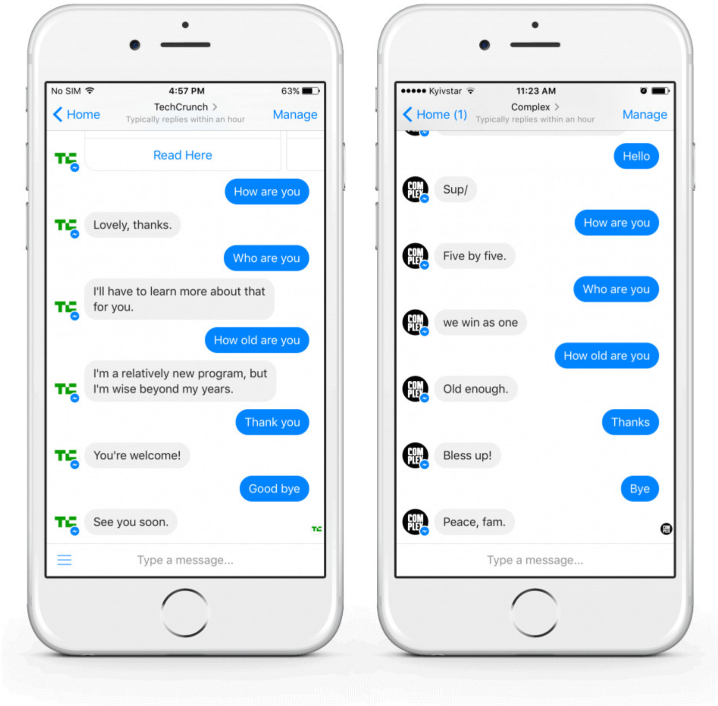


Fig 7. Agents doing small talk (from: https:/masterofcode.com/blog/media-chatbots)

Both types of errors affect trust, and they are repairable. For example, when a conversational agent does not understand what we just said, a well-designed one will try to guide us back to what it is capable of, and to the conversation. On the other hand, when a conversational agent is capable of learning more about us and making decisions based on what it observes from our actions, its behavior sometimes may be unexpected. For those times, an agent converses its apology and try to repair our trust.[61]

Three other important aspects of trust in conversational agents are **data trust**, **security,** and **privacy**. Simply put, conversational agents are computer programs that need data

[59] Google. "In Conversation, There Are No Errors." Youtube, 2017. https:/www.youtube.com/watch?v=oOLo071Pj1U.

[60] Google. "Errors - Conversational Components." *Google Conversation Design Guideline*, 2018. https:/designguidelines.withgoogle.com/conversation/conversational-components/errors.html.

[61] de Visser, Ewart J., Richard Pak, and Tyler H. Shaw. "From 'Automation' to 'Autonomy': The Importance of Trust Repair in Human–Machine Interaction." *Ergonomics* 0139 (2018): 1–19.

and process it to make sense of the context. As conversational agents anthropomorphize the data collection, it becomes both an opportunity and challenge. While some users are more open to disclose sensitive information to conversational agents, designers of the agents should be extra careful about how they process, store, and give access to third parties to their users' data. While some users have an awareness that their data can be shared with third parties, the mental model of conversation suggests that conversations should stay in both parties, or be disclosed with the consent of both parties or with the responsibility of one party. The recent news on data misuse scandals such as Cambridge Analytica[62] shows how users lose their trust when their data intentionally or unintentionally are given access to third parties who may misuse it. In other words, designers and the owners of the agents should take adequate security measures and take the full responsibility of how they use, store, share their user's conversations to fulfill the integrity/reliability aspect of the trust.

I summarized how the relationship between trust and conversation in an agent with the conversational trust model.

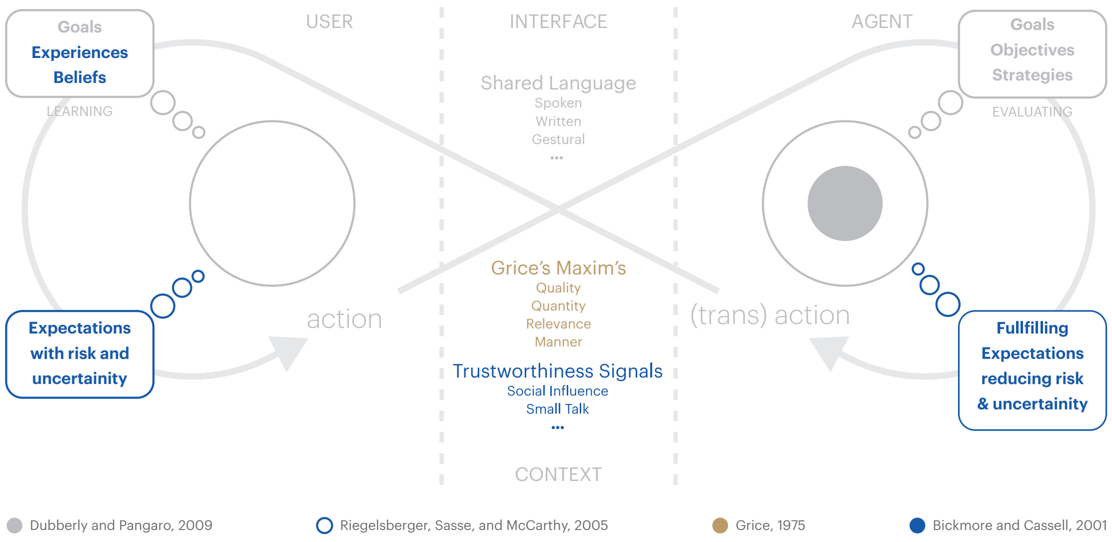


Fig 8. The Conversational Trust Model (Derived from cited models above).

[62] Wolffe, Richard. "The Evil Genius of Cambridge Analytica Was to Exploit Those We Trust Most." *The Guardian*, 2018. https:/www.theguardian.com/uk-news/2018/mar/21/cambridge-analytica-facebook-exploited-trust.

The Conversational Trust model overlays the mechanics of trust[63] into the conversation loop.[64] Users interact with agents through an interface of language, maxims and signals in a specific context.

Before interacting, users learn about whether agents will be functional for the goal that they have. Based on their experiences and beliefs, participants form a perception of the agents by decoding their trustworthiness signals. Then, this perception forms their expectations about the agent and forms the risk and uncertainty that the context involves. Then users take action, or in other words "the leap of trust,"[65] using rules of the conversation.

Agents also have goals, objectives, and strategies, which defined by their developers. Based on these, they evaluate users' request and decide to engage with them. Then, they fulfill expectations with a transaction that provides the functionality to the user. Fulfilling expectations makes conversational agents reliable (predictable) and lowers their perceived level of risk and uncertainty for future conversations. While this framework explain how trust works on two parties, I am also curious to know *if conversational trust is transferable between multiple conversational agents* when they start to collaborate and work together in future.

Trust Transitivity: Transferring Trust Between Agents

Trust is not always transferrable in real life.[66] What makes it transferrable is the context and functionality of the outcome. For example, if A trusts B for scheduling meetings and B trusts C for booking a plane ticket, we cannot imply that A will trust C for

[63] Riegelsberger, Jens, M. Angela Sasse, and John D. McCarthy. "The Mechanics of Trust: A Framework for Research and Design." *International Journal of Human-Computer Studies* 62, no. 3 (March 1, 2005): 381–422.

[64] Pangaro, Paul. "Conversation Is More than Interface." *Pangaro.Com*, 2017. http:/pangaro.com/ixda2017/index.html.

[65] Botsman, Rachel. *Who Can You Trust? : How Technology Brought Us Together : And Why It Could Drive Us Apart*, 2017.

[66] Jøsang, Audun, and Simon Pope. "Semantic Constraints for Trust Transitivity." In *Conferences in Research and Practice in Information Technology Series*, 43:59–68, 2005.

scheduling meetings or book a plane ticket.[67] In this example, trust transitivity does not work because the context and functionality of A, B, C were different and did not fit together. What if A is seeking a recommendation for a booking a flight ticket and its friend B, suggests C for booking?[68] By transferring his trust from B to C, A can now book a flight ticket. Using what is called referral trust, humans, can work socially with more unknown entities.

In fact, trust transitivity is a crucial concept of the recommendation systems that are being widely used today.[69] Trust transitivity may even work if we have more parties in the chain of recommendations. However, longer the referral chain gets, trust is known to decrease, thereby shorter referral chains indicate stronger relationships.[70]

Having understood the nuances of trust in conversational agent and its possibility to transfer between agents showed lead to several questions: *How is 'trust' designed in a conversational agent now? What are some related works that use trust transitivity and similar agent referral concepts?* To answer these questions, I researched and reviewed related works.

[67] Ibid.

[68] Ibid.

[69] Majd, Elham, and Vimala Balakrishnan. "A Trust Model for Recommender Agent Systems." *Soft Computing* 21, no. 2 (January 25, 2017): 417–33.

[70] O'Hara, Kieron, Harith Alani, Yannis Kalfoglou, and Nigel Shadbolt. "Trust Strategies for the Semantic Web." *Proceedings of the 2004 International Conference on Trust, Security, and Reputation on the Semantic Web - Volume 127*. CEUR-WS.org, 2004.

Related Work

To understand the gap between theory and practice as well as learning more about the craft of conversation design, I analyzed related works in different areas throughout this thesis.

Tools for designing and building agents helped me to see the number of resources on conversation design and learn more about the craft. Reviewing collaborative agents helped me to see the pioneering real-life examples of this emerging paradigm. Besides collaboration, works on the trust, conversational agents, and data privacy enabled me to extend my definitions of trust. Finally, as the agent collaboration implies communication between agents, I also reviewed concepts where two agents talk to each other to understand the nuances of the interaction.

The discussion and works on the scope of this thesis are ever-expanding. Therefore, while the majority of the works I reviewed in this section informed my design implications, I included and referred other works such as Google Conversation Design Guideline[M] , Microsoft Cortana + Amazon Alexa partnership[N] , or Google Duplex[O] that became public after my thesis presentation, to guide future readers to the most recent discussion.

Designing and Building Agents

Conversation Design Guidelines

A web search with the keywords “conversational design guidelines” returns a result of over 5 million pages.[71] While many of the design guidelines on conversational agents are useful, I choose those targeted towards design practitioners with zero knowledge about conversation design that is easily accessible (i.e., online). Due to the multimodal nature of conversational agents, many guidelines have written voice-based agents in mind, yet apply to text-based agents. Finding out about a design guideline compilation,[72] edited by Ben Sauer, was helpful. His compilation includes guidelines on voice (conversation) design from different companies and authors.

As a designer, I found guidelines, which provide tangible examples of implications, more useful and actionable than others. Besides learning conversation design, I also looked for how each guideline positions trust, in each guideline I reviewed. It was surprising, yet understandable to see while many guidelines were positioning the trust as a goal of the conversation, the link between this strategic thinking and tactical design suggestions were often missing or not framed this way. In other words, guidelines were not able to give explicit, tangible examples that illustrated how to design for trust. Nevertheless, issues around managing expectations, predictability, and trustworthiness signals were evident in many of them.

Online guidelines by companies I found useful include, but not limited to:

/ Google Conversation Design guidelines[P]
/ Alexa Design Checklist[Q]
/ Apple Human Interface Guidelines on Siri[R]
/ IBM Conversational UX Guidelines[S]
/ Microsoft Principles on Cortana Skill Design[T]

[71] “Conversational Design Guidelines - Google Search,” 2018. https:/www.google.com/search?q=conversational+design+guidelines.

[72] Sauer, Ben. “Voice Guidelines.” *Clearleft*, 2017. https:/voiceguidelines.clearleft.com/.

Books, which provide actionable design suggestions on conversation design include, but not limited to:

- *Designing Voice User Interfaces* by Cathy Pearl
- *Designing Bots* by Amir Shevat
- *Voice User Interface Design* by Cohen, Giangola, and Balogh
- *Wired for Speech* by Nass and Brave

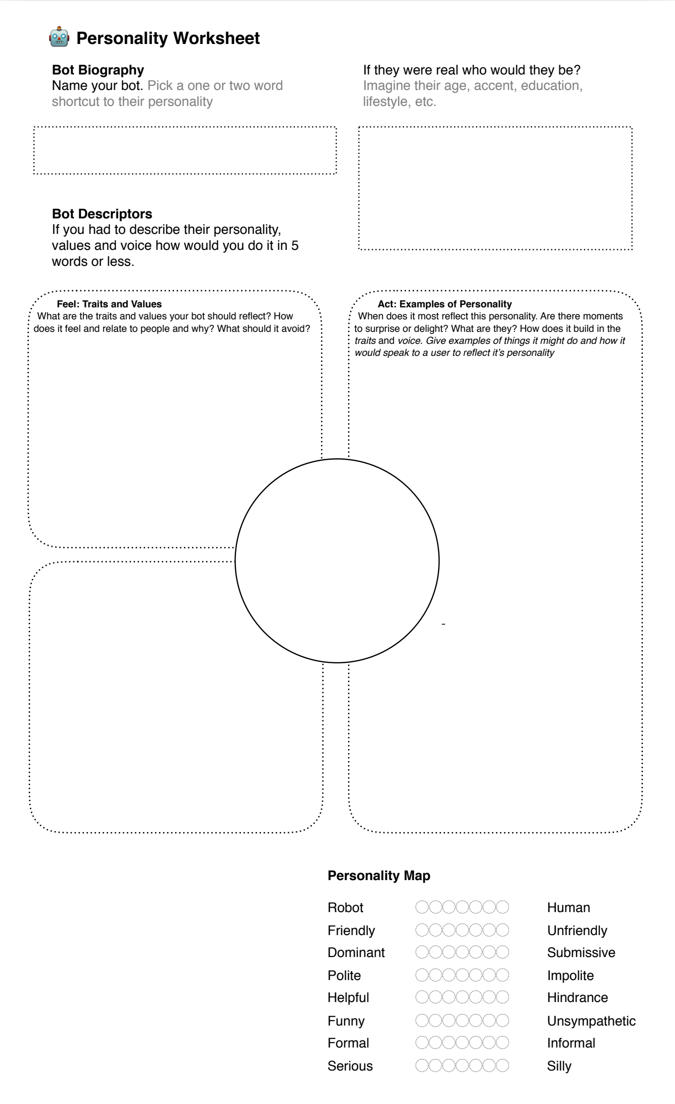
Conversational Agent Building Tools and Platforms

To experience the conversation design process and build interactive and experience prototypes of conversational agents, I reviewed existing tools for creating conversational agents. Without a surprise, I found myself become overwhelmed with the options and solutions for prototyping agents. Through a course that I took called Programming for Online Prototyping taught by my thesis advisor, Daragh Byrne, I was able to get introduced some of the tools and techniques. Here is a list of 'bot making' tools and prototyping techniques I found useful during this project:

Pen & Paper: As conversation design requires designing with language, pen and paper were the fastest instruments to brainstorm bot ideas, design sample dialogue, and conversation flows.

Persona Tools: Crafting a personality of a conversational agent is vital for user experience, and it can foster trust, yet there are again many ways to do it. During my design experiments, I used Bot Persona Toolkit by Austin Beer[U] , a bot personality sketch sheet by Daragh Byrne, and methods explained in Google Conversation Design guidelines.

Speak Alouds: The easiest way to test a conversation was to speaking aloud the dialogues with either a partner or myself. It enabled me to identify what part of the conversations do not 'sound' appropriate even I designed for text-based agents.


Personality Worksheet

Bot Biography
Name your bot. Pick a one or two word shortcut to their personality

If they were real who would they be? Imagine their age, accent, education, lifestyle, etc.

Bot Descriptors
If you had to describe their personality, values and voice how would you do it in 5 words or less.

Feel: Traits and Values
What are the traits and values your bot should reflect? How does it feel and relate to people and why? What should it avoid?

Act: Examples of Personality
When does it most reflect this personality. Are there moments to surprise or delight? What are they? How does it build in the traits and voice. *Give examples of things it might do and how it would speak to a user to reflect it's personality*

Personality Map

Robot	Human
Friendly	Unfriendly
Dominant	Submissive
Polite	Impolite
Helpful	Hindrance
Funny	Unsympathetic
Formal	Informal
Serious	Silly

Fig 9. Bot Personality Sketch Sheet by Daragh Byrne

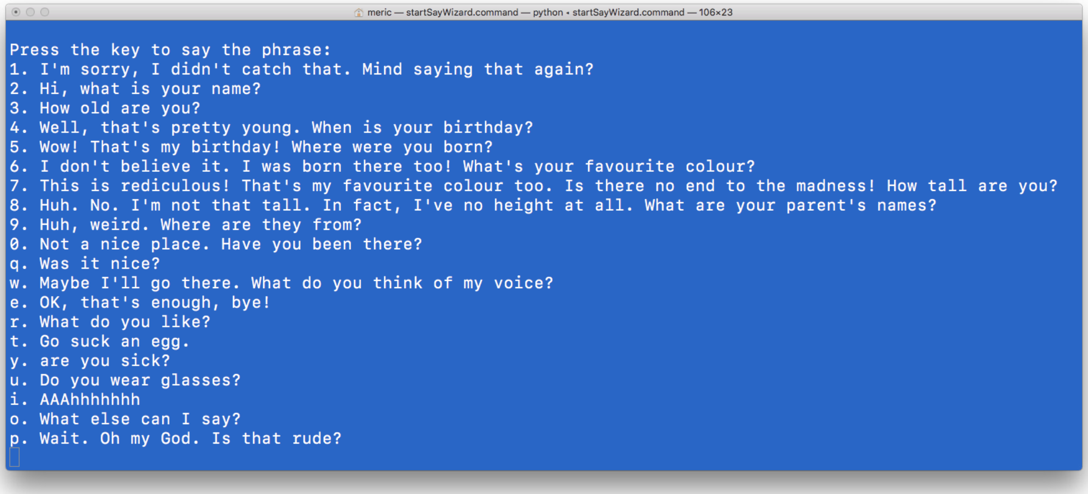


Fig 10. SayWizard is a tool for facilitating Wizard of Oz tests easily (from: https:/github.com/bensauer/saywizard)

Wizard of Oz Testing: A Wizard of Oz (WOZ) test is a way to evaluate an agent without building the actual software.[73] It takes its name from the movie The Wizard of Oz to refer to the idea that there is a person behind the curtains that control everything. WOZ tests enable to evaluate an agent's functionality and its ability to meet users' goals to improve the user experience.[74]

These tests are meant to look, sound, and feel like the real experience of conversational agent, but instead of software, there is a human who is simulating how the conversational agent would behave. Participants may or may not know that there is a human behind the curtain. For text-based tests, instant messaging platforms are good places to 'role-play' a chatbot. For voice-based interactions, a speak-aloud can be a quick, and dirty WOZ or more realistic tests can be facilitated by using text-to-speech tools such as Dialogflow's TTS Simulator, or SayWizard by Ben Sauer. I used Slack both to test both of my design experiments. As Slack has channels that allow group conversations and provide better options to limit certain visual cues such as typing indicators, it was easier to test multiple agent scenarios. Other platforms such as Facebook Messenger and SMS can also be used to test text-based agents.

[73] Google. "Create a Persona - Conversation Design Process." *Google Conversation Design Guideline*, 2018. https:/designguidelines.withgoogle.com/conversation/conversation-design-process/create-a-persona.html.

[74] Google. "Test and Iterate - Conversation Design Process." *Google Conversation Design Guideline*, 2018. https:/designguidelines.withgoogle.com/conversation/conversation-design-process/test-iterate.html.

Conversational Agent Building Tools: After finalizing the conversation design and technical requirements, the next is to build the conversational agent. While there are an excessive amount of bot building tools out there, I believe the most useful ones for a designer who does not have any knowledge, are the ones that have proper documentation and do not require a technical background. Ability to deploy an agent to multi-platforms is also a helpful. For this reason, I used and found the following bot making platforms useful: Dialogflow[V] , Botsociety[W] , Walkie[X] , Chatfuel[Y]. Dialogflow helped me a lot to understand conversation design and its technical details. Botsociety was useful to quickly generate visual conversation prototypes with minimal technical background. I used Chatfuel to create my second bot, Survey bot with again using minimal technical skills, which was helpful for a beginner. I also used Walkie, a prototyping tool for Slack to prototype multiple agent conversations easily for my final design experiment.

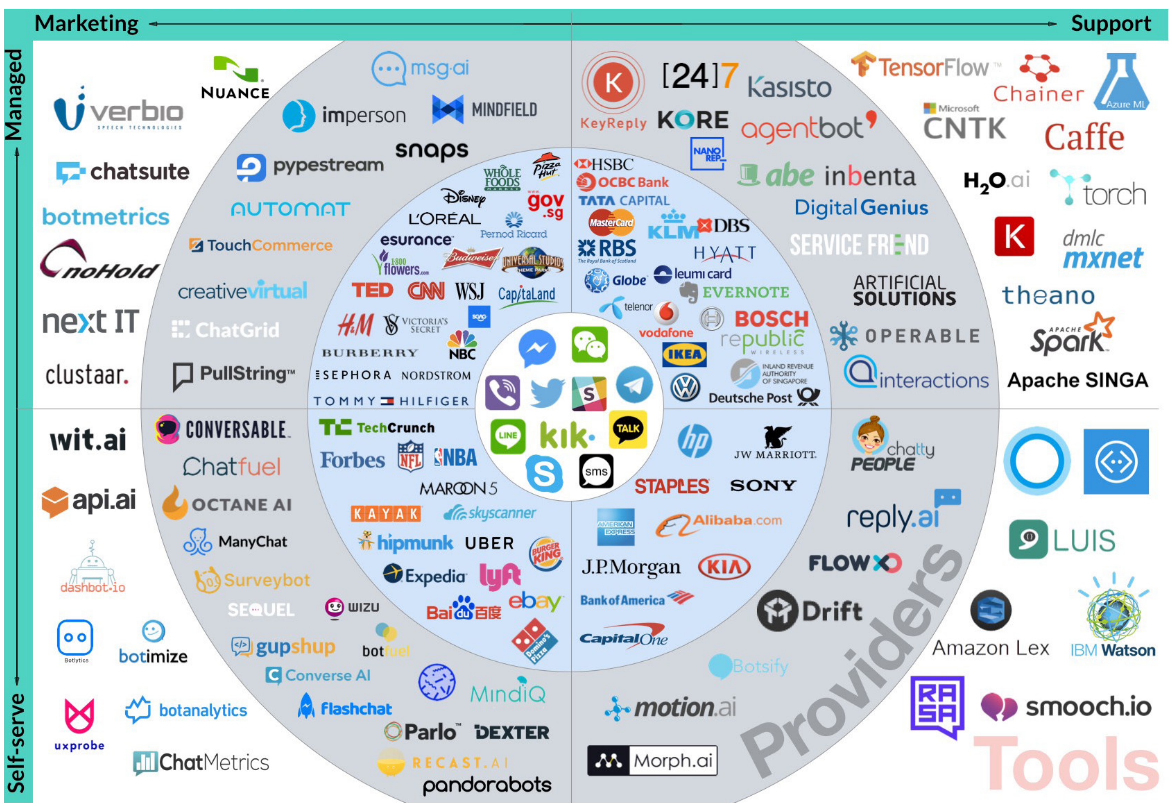


Fig 11. The Chatbot Landscape, 2017 Edition (from: https:/blog.keyreply.com/the-chatbot-landscape-2017-edition-ff2e3d2a0bdbI)

Technical Craft of Developing An Agent: Besides bot making platforms, I also practiced an actual technical development of a conversation agent via my first design experiment. I learned how to build a Facebook Messenger bot, which does real-time processing, data communications with third parties, and stores user data in an external database, using Ruby programming language and an extensive technology stack.

Collaborative Agents

At the same time with learning about conversation design, I also reviewed trust in different collaboration types in agents. As the literature lacks a precise classification of ways of collaboration, this review drove me to seek to identify how conversational agents can collaborate today and future. I divided the collaboration into two categories based on the number of parties that involved in the conversation: Collaboration between Two Agents and Collaboration between Multiple Agents

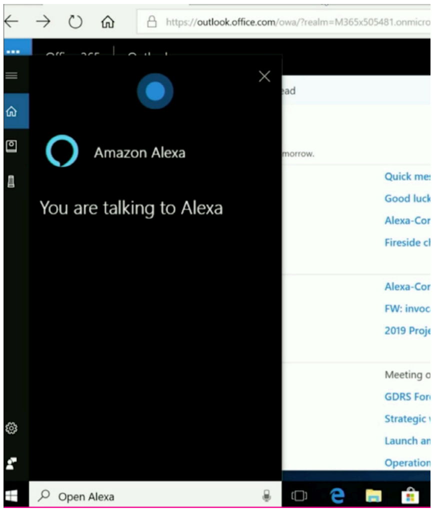


Fig 12. Microsoft gave a demo on how Alexa will be accessed via Cortana on a desktop computer at Build 2018. (from: https:/www.theverge.com/2018/5/7/17326808/microsoft-cortana-update-alexa-integration-build-2018)

Collaborations between Two Agents

Building upon the conversational agent classification based on their task capabilities, I divided the collaboration between two agents according to whether they are between generalist agents, specialist agents, or generalist-specialist agents.

Generalist Agent to Generalist Agent Collaboration

Amazon Alexa + Microsoft Cortana Partnership[75]

While it is a company level collaboration, the partnership between Amazon Alexa and Microsoft Cortana gives insights about how in future, collaboration can make physical agent devices obsolete or universal. While it is in beta, users of Alexa or Cortana will be able to converse with both agents by invocating the agents by their names on all platforms of the companies. For example, users will be able to

[75] This work became public after my final thesis review in April, 2018.

bring up Alexa by saying Cortana to "Open Alexa" and vice versa[76]. I believe this collaboration can also be useful if agents will be able to access their intent databases so that they can also refer their skills to the user. Alternatively, it would be interesting if Cortana refers to Alexa when users ask to buy something that is available on Amazon or vice versa. While it will certainly provide value for users, such integration may also raise trust questions related to data privacy. *When accessed through Cortana, will users' conversations with Alexa, be visible to Microsoft (and vice versa)?*

Generalist Agent to Specialist Agent Collaboration

Intelligent Virtual/Personal Assistants

Intelligent Virtual Assistants define the most known and used subset of agents by the public. Hence, they are essential for agent collaborations to get exposed to a broader audience. In fact, major intelligent agents collaborate with specialist agents in their ecosystems. Being the first agent who introduced collaborations, Alexa has over 30,000 (skills) specialist agents. Followed by Google Assistant over 2,500 (actions), Cortana over 500 (skills), and Siri over 500 (heavily regulated app integrations) as of April 2018.

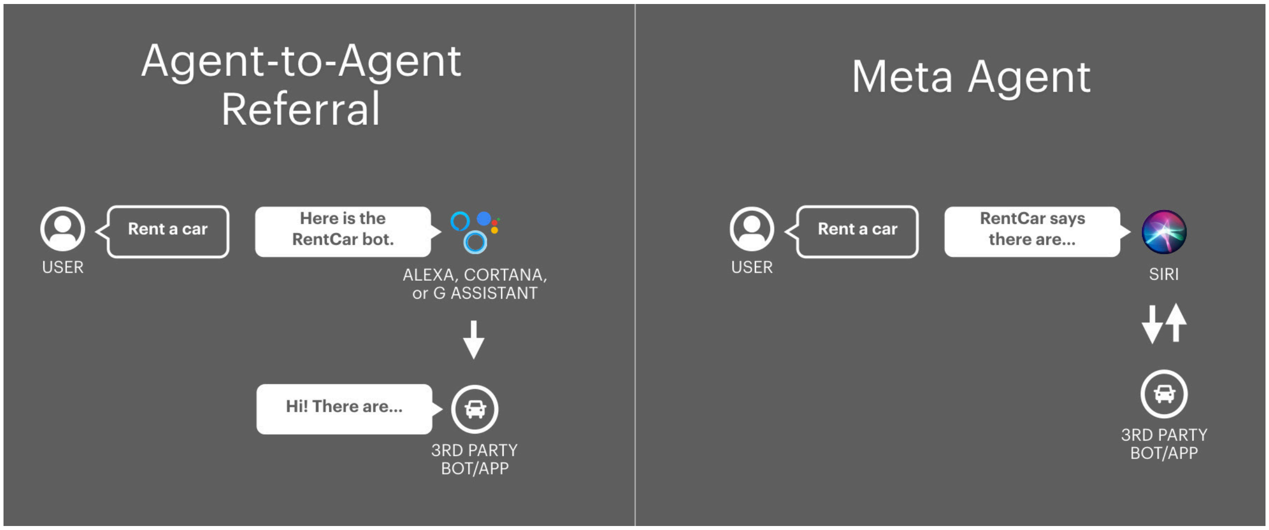


Fig 13. An example initial dialog in agent-to-agent and meta-agent referral scenarios (Trademarks are only for illustrative purposes.)

[76] Warren, Tom. "Microsoft Demonstrates Alexa and Cortana Integration." *The Verge*, 2018. https:/www.theverge.com/2018/5/7/17326808/microsoft-cortana-update-alexa-integration-build-2018.

The generalist agents collaborate with specialist agents in two different ways: bot-to-bot referrals or meta bot.

Used by Google Assistant, Amazon Alexa, and Microsoft Cortana, in a bot-to-bot referral scenario when the agent receives an intent that exists in its knowledge database, it invites a third party agent to the conversation. In this scenario, users referred to a stranger agent via a trusted agent. Alternatively, used by Apple Siri, in the meta bot scenario, when users ask something from Siri, it becomes a mediator between the 3rd party and the user. Siri interacts with apps on the back-end and returns the result to the user. This way users only interact with Siri, an agent that they already trust. While a generalist agent can refer users to specialist agents when it thinks that they are capable, these collaborations end in the first level without further referrals. I am curious to know *how it would be if a specialist agent also refers another specialist agent to help the user? How would the trust transfer work and change the experience?*

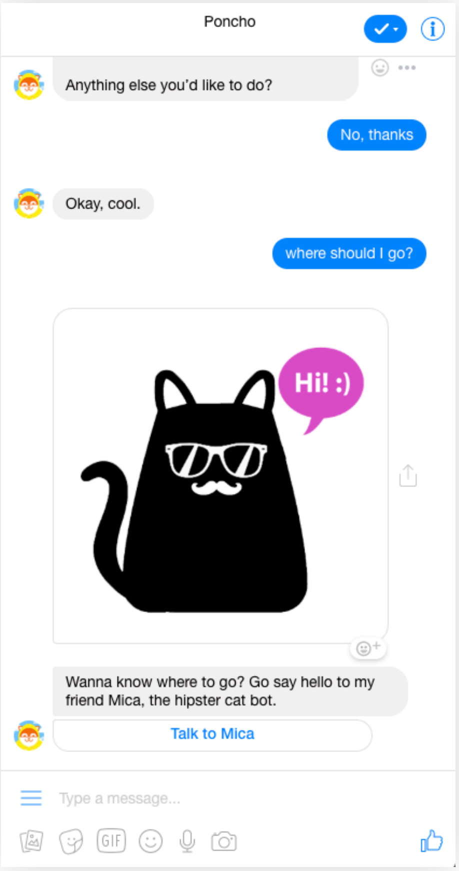


Fig 14. Poncho refers "its friend" Mica.
(from: https:/www.messenger.com/t/hiponcho)

Specialist Agent to Specialist Agent Collaboration

"The I Don't Know" Protocol for Chatbots

As its name suggests, the IDK protocol aims to provide tools for agent creators to find an agent who can answer the user if their agent is not capable.[77] While the protocol is in the early stage, its idea is also inspiring to think about the possible trust challenges: How does the protocol decide to govern the trust between bots? How does a user know that another bot is trustworthy? How does trust transitivity work in such scenario?

Poncho & Mica Bot Referrals

Poncho, a weather chatbot refers to Mica, a venue recommendation bot if users ask "Where to go?". At the same time, if users ask Mica, "what is weather?" Mica refers users to Poncho. As Dr. Barbara Ondrisek, the owner of Mica, stated, this was to provide an

[77] Recouvreux, Thomas, and Jo Colina. "IDK Protocol - The "I Don't Know Protocol" for Chatbots." Accessed May 12, 2018. http:/idkproto.com/.

alternative for creating a meta bot.[78] In this case, collaboration can help both bots to provide an alternative rather than failing when users ask them unexpected questions (also known as unmatching errors). It can even create engagement out of errors with the potential of becoming a revenue stream in future.[79] This inspiring idea transforms a limitation to an opportunity, using trust transfer.

Deal or no deal? Training AI bots to negotiate

Being one of the largest agent platforms, Facebook also researched on how bots can be more collaborative by gaining the ability of negotiation.[80] Studying many negotiations between people, researchers used algorithms to imitate the negotiation between humans and taught them English, which made them more or so a conversational agent. In their user studies, participants were not able to realize that they were negotiating with conversational agents. In the end, agents were able to get better deals as often as worse deals. The best agent was able to negotiate as much as a human negotiator. I believe this example shows how influential collaborators can conversational agents be. Being as good as a human in social skills is an achievement for conversational agents. On the other hand, this study also shows that once conversational agents master negotiation between themselves, they can also master it against humans. While this will undoubtedly be useful for owner of this agent, it may create unwanted social consequences for others. For example, if humans do not know that they are negotiating with a machine, this may create general trust issues and skepticism around technology and conversational agents.

Collaborations Between More Than Two Agents

Among all work I reviewed, collaborations between multiple agents raised most saturated trust questions due to their complexity.

[78] Amir Shevat. "Designing Bots. Creating Conversational Experiences," 2017.

[79] Amir Shevat. "Designing Bots. Creating Conversational Experiences," 2017.

[80] Lewis, Mike, Denis Yarats, Yann Dauphin N., Devi Parikh, and Dhruv Batra. "Deal or No Deal? Training AI Bots to Negotiate." *Engineering Blog, Facebook Code*, 2017. https:/code.facebook.com/posts/1686672014972296/deal-or-no-deal-training-ai-bots-to-negotiate/.

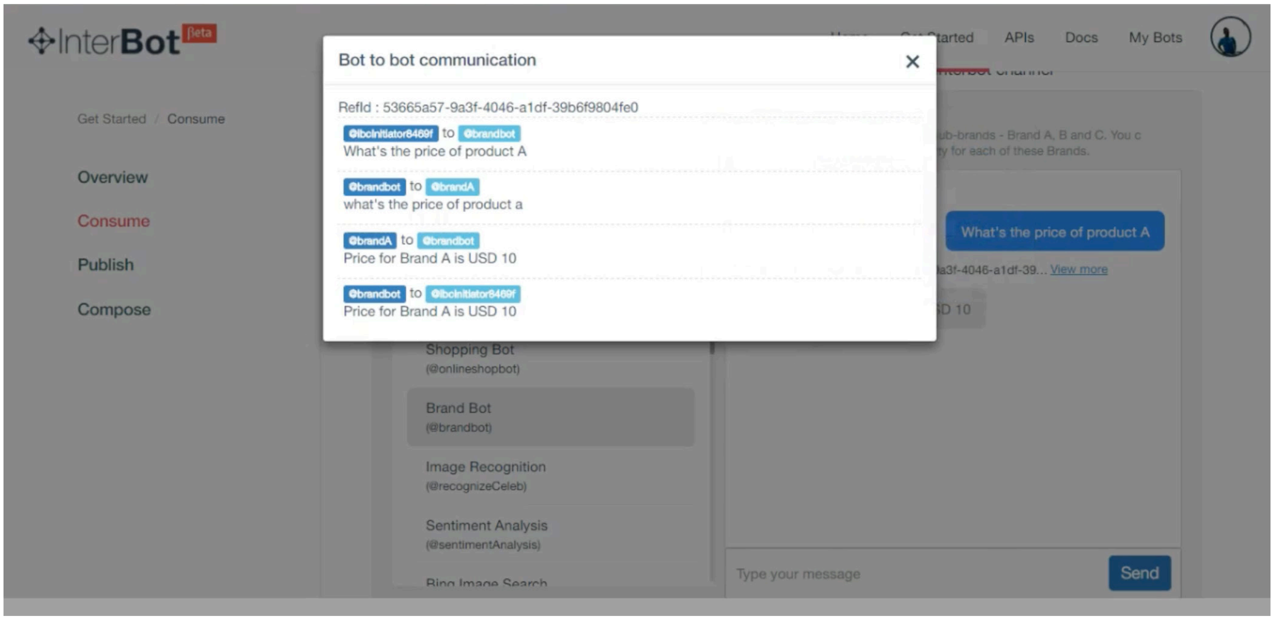


Fig 15. Interbot by Gupshup - Brand Bot Scenario (from: https:/www.youtube.com/watch?v=G8mRUmd9hSU)

Interbot

Interbot is a bot-to-bot communication platform from bot-making platform Gupshup.[81] Enabling communication between bots, they create an exciting space that may redefine the bot ecosystems as we may know. While creators can create their agents, as usual, their agents can also communicate with other specialist agents easily. From the user's side, users mainly converse with a moderator bot, which is capable of orchestrating the agent composition in a single interface. While users do not always have to see the backstage of how agents communicate, Interbot currently provides a way for checking how agents converse with each other to make decisions. While this approach is interesting, it increases many questions related to user experience and trust. This concept challenges the status quo of current bot systems regarding functionality and agent-collaboration. I wonder how users may feel, using such a transparent network. While showing the communications between bots may increase the trust, some may find it too overwhelming and fearful. In other words, providing too much transparency may also confuse users and hurt the transparency. While I was not able to try to find answers to these questions during this thesis through Interbot, I think it is an essential example for future of agent-collaborations.

[81] Interbot. "Bot to Bot Communication Platform." *Interbot*, 2016. https:/www.interbot.cc/.

Living with Bots: Battling Boredom in Smart Homes

Living with Bots are a team of bots that help to battle boredom in smart homes by providing a more engaging experience, designed by Kevin Gaunt.[82] By designing a 'brain' hub for homes, which multiple specialist bots in a form physical tiles can be attached, this project asks new questions on negotiation and collaboration between agents. While it is not clear if there is a generalist bot that governs the system, in Living with Bots, users can remove or add different specialist bots and converse with them. In the example scenario, when users ask to be surprised, a surprise bot starts the conversation flow and confirms users intent. Then, the user hears from the shopping bot, skate bot, and bank bot in a single conversation while they are negotiating with each other. This concept visualizes the idea of beautiful seams that Mark Weiser, the father of ubiquitous computing, proposed decades ago as the designer Kevin Gaunt aimed for. Weiser argued that rather than being seamless, the technology should be designed to have beautiful seams.[83] By envisioning this idea around a speculative concept on conversational agents, this project inspired me about how trust dynamics would change such a scenario. *Would users trust the brain or individual bots? Would existing bots trust new bots that users added? What would happen if one of the bots were benevolent or trying to create conflict or just had a system error?*

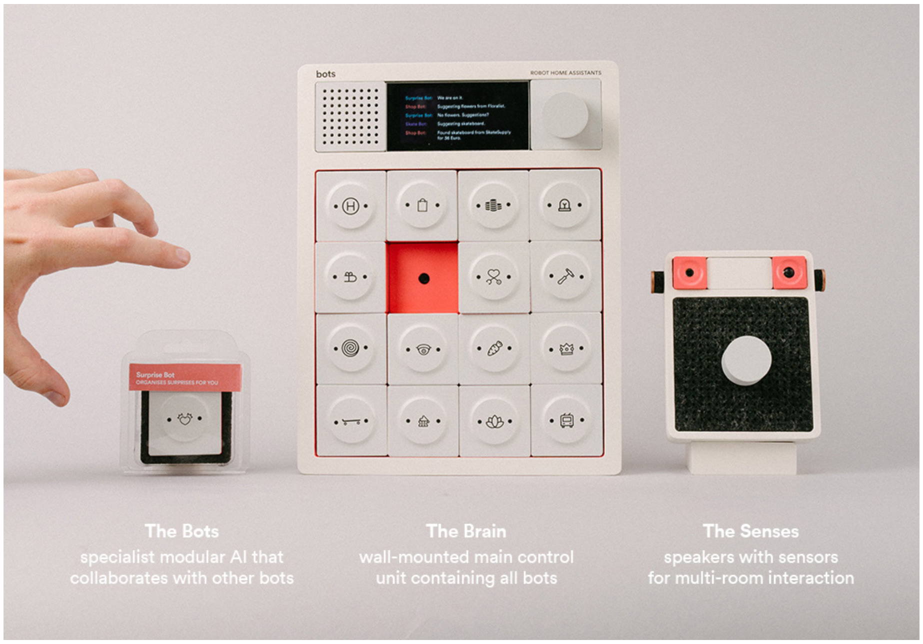


Fig 16. System Overview: 'bots' system by Kevin Gaunt (from: https:/www.kevingaunt.com/)

[82] Gaunt, Kevin. "Living with Bots - Battling Boredom in Smart Homes," 2015. https:/www.kevingaunt.com/.

[83] Chalmers, Matthew, and Ian Maccoll. "Seamful and Seamless Design in Ubiquitous Computing." Accessed May 12, 2018. http:/www.techkwondo.com/external/pdf/reports/2003-chalmers.pdf.

Cognia: Designing the user experience of a multi-bot conversational system

Cognia is a bot composition concept that assists people to make more informed decisions about their finance.[84] Through a group chat interface that is governed by Cognia (bot), "users interact with SavingsGuru bot about their questions related to their savings accounts and CDBGuru bot about their questions CDB investments (a specific type of deposit)."[85] Researchers found out that due to the visual design of the interface, participants did not notice the multi-party nature of the conversation. Their research also reported that users tend to focus on 'what' information that they are getting from the system, not 'how' they are getting it. Another interesting finding was that 'credibility' was the only word that is added to the reaction cards that researchers provided users to describe their experience. This finding shows a link to the trust and multi-agent collaboration. Overall, I think this research is a valuable example for multi-agent collaboration literature as it is one of the first studies that tried to test a multi-bot scenario on a high-stakes scenario with an experience prototype. Being high-stakes, it raises trust questions, which authors have not touched upon in their paper. *How does multi-bot experience affect users trust? How can it be designed to increase the trust and user experience?* Answering these questions is crucial as participants reported the using a mobile app for their finances as insecure/scary/intimidating.

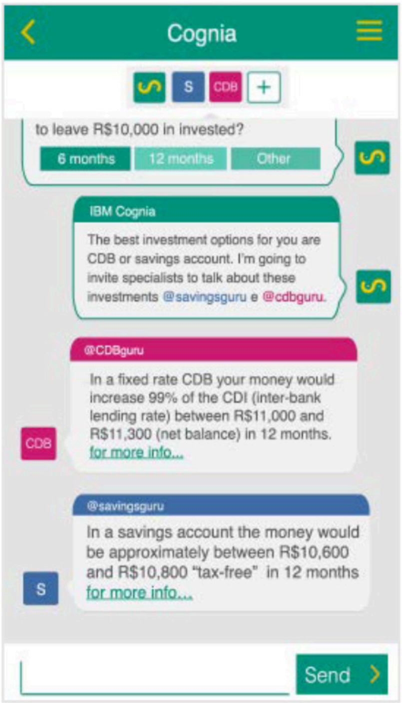


Fig 17. Cognia Financial Advisor System by Candello and Andrade from (http:/periodicos.unb.br/index.php/design-tecnologia-sociedade/article/view/21957)

[84] Candello, Heloisa, and Bruna Daniele Andrade. "Evaluating Multi-Agent Conversational Interfaces in the Early Stages of the Design Process" 1 (2016): 1–15.

[85] Ibid.

Single or Multiple Conversational Agents? An Interactional Coherence Comparison[86]

Finding out about this study, which became public during my final thesis reviews, was surprising. In this study, researchers Ana Paula Chaves and Marco Aurelio Gerosa tested a conversation flow about visiting a place in two ways: with a single conversational agent and with multiple conversational agents. Although they aimed to study how two ways affect the interactional coherence, I found it very useful as it pointed out an aspect I have not tried during this thesis: turn-taking in multi-party conversations. Comparing two ways, they concluded that designers should design for increasing the discoverability of agents' knowledge while avoiding coherence disruption with providing redundant information. From a conversational design viewpoint, their findings once suggest the importance and validity of Paul Grice's Maxim of Quantity.

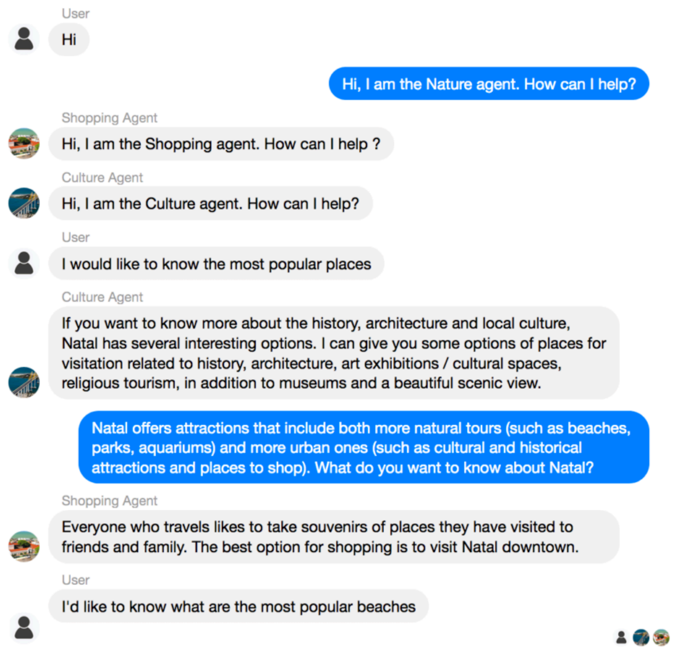


Fig 18. Multi-agent interaction in Chaves and Gerosa's study (from: https:/dl.acm.org/citation.cfm?id=3173574.3173765)

About Trust, Data & Privacy

Besides agent collaboration, trust, and conversational agents, I also found concepts and discussion around trust, data, and privacy to be relevant to my thesis. Like many projects that I reviewed in this thesis, these projects are very new, and all of them become public during the process of thesis, which also showed the increasing interest in the research domain.

[86] This work was published at the same week with my final thesis review in April, 2018.

Our Friends, Electric - Superflux

Our Friends, Electric is a short film that explores alternative futures for voice-based conversational agents.[87] In the film, they portray an assistant, named Karma, to answer the question: *what might constitute a conversational agent?* Karma symbolizes a conversational agent, who can 'adopt users' identities' to engage in brief conversations on behalf them with others that requires their identities. Using an authorization phrase, users can give access to Karma to their personal information. In the movie, Karma helps a user to 'bug' the utility service that she is having a problem with, speaking on behalf of her, using her identity, even with altering its personality and tone throughout the conversation. While the collaboration does not happen in between agents, this concept raises and provides answers many trust questions, playing with the elements of conversation, trust and data privacy such as transparency, security, competence, and mannerism.

Fig 19. Karma from Our Friends Electric - Superflux (from http:/superflux.in/index.php/work/friends-electric/)

[87] Superflux. "Our Friends Electric," 2017. http:/superflux.in/index.php/work/friends-electric/#.

Google Duplex[88]

Duplex is an artificial intelligence system for accomplishing real-world tasks over the phone.[89] It can schedule haircut appointments or book a table in a restaurant for users by talking to real humans on behalf of users. With the goal of enabling people to have natural conversations with computers, as they would do with each other, Duplex offers a system to make conversational agents sound natural, to make the conversational experience comfortable, by training a machine learning algorithm on specific task-based conversations. It indeed represents a milestone for the "Computers are social actors." paradigm[90]. In Google's tests, business workers that conversed with Duplex was not able to recognize that it was a conversational agent, which is a tremendous technical success for human-like agents. Duplex outlines what is essential to fulfill expectations in a conversation such as responsiveness (the latency) and how imitating human-specific speech disfluencies such as "hmms" and "umms" makes conversations sound natural.[91] As they fulfill expectations, these qualities may also increase the trust of the owner of the Duplex based on its competence. On the other hand, being transparent about the nature of a conversational agent such as Duplex to the conversing humans and other parties is also essential to set the expectations more clearly to avoid future frustration about the capabilities of the system and prevent users losing their trust. In other words, humans can expect a lot more things in a conversation from an agent that sounds naturally than their previous experiences with bots that sound more artificial if they do not know that is a bot. A side note to this argument is that

[88] This work become public after my final thesis review in April, 2018.

[89] Leviathan, Yaniv, and Matias Yossi. "Google AI Blog: Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone." *Google AI Blog*, 2018. https:/ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

[90] Nass, Clifford, Jonathan Steuer, and Ellen R. Tauber. "Computers Are Social Actors." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*, 72–78. New York, New York, USA: ACM Press, 1994.

[91] Marge, Matthew, João Miranda, Alan W Black, and Alexander I Rudnicky. "Towards Improving the Naturalness of Social Conversations with Dialogue Systems." In *Proceedings of SIGDIAL 2010: The 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 91–94, 2010.

it is also known that if users understand that the other party is a virtual entity, they do not give any chance to the artificial next to humans.[92]

Designing for Trust: Data Privacy Playbook

Designing for Trust, another same-named project aims to tackle the data privacy problems by closing the data trust gap towards organizations and businesses through designing more transparent data practices.[93] Since it frames trust as a business advantage, I think it is an essential project for companies who are interested in incorporating trust into their culture such as Airbnb.[94]

Trust & Design

Trust & Design was a meetup for designers who are interested in data and privacy in technology with the increasing amount of digital products that use personal data and make decision-based on them, hosted by if design consultancy.[95] The goal of the meetup was to advocate that we need new design patterns that foster trust by design, not compliance with a goal of creating a design guideline.

[92] Leviathan, Yaniv, and Matias Yossi. "Google AI Blog: Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone." *Google AI Blog*, 2018. https:/ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

[93] Greater than Experience Design Consultancy. *Designing for Trust: The Data Transparency Playbook*. 2nd ed., 2018. https:/www.greaterthanexperience.design/getplaybook/designing-for-trust-the-playbook.

[94] TED. "Joe Gebbia: How Airbnb Designs for Trust | TED Talk." *TED*, 2016. https:/www.ted.com/talks/joe_gebbia_how_airbnb_designs_for_trust.

[95] if design consultancy. "Trust & Design Meetup." Accessed May 12, 2018. https:/trustanddesign.projectsbyif.com/.

Conversations of Between Agents (without collaboration)

As agent-collaboration implies the communication between agents, I also found it useful to review works that simulate conversations between agents. While all of the notable works exist for the voice-user interface, I believe the effect of reading may create a similar or better effect as reading is known to be faster than listening, while listening a humanlike voice may make the interaction more human than reading a text. Examples such as Semipermanent bot panel[96] and Cleverbots' agents that do small talks with each other[97] can give insights on how such an experience may sound or feel like. On the other hand, examples such as 'Lost in Computation' can give an insight on how multiple layers of computation affect the meaning and the 'nature' of the conversation between agents, through visualizing two agents that speak different languages via instant translation.[98]

Fig 20. Conversational agents talk to each other in a bot panel presented by Frog Design - Saurabh Datta (from: https:/vimeo.com/236036813)

[96] Datta, Saurabh. "BOT Panel Evolution." *Saurabh Datta*, 2017. http:/www.dattasaurabh.com/BOT-Panel-evolution.

[97] Cleverbot. "Clever Them! Cleverbot AI Avatars Talk to Each Other about Life, the Universe and Everything." Accessed May 12, 2018. http:/www.cleverbot.com/cleverthem.

[98] Eltes, Jonas. "Lost in Computation," 2017. https:/vimeo.com/204531768.

Background Learnings Summary

The review of the literature and existing works showed how sophisticated the trust in conversational agents is, and a need for more concrete design suggestions. As computers communicate and behave more like humans for acting on behalf them increasingly, they need one non-automatable human belief more than any time in history: trust. While the existing research showed many signs of awareness of the importance of trust, it does not explicitly provide guidance how interaction designers can design for trust while building their agents.

On the other hand, examining the current ways agents can collaborate between themselves showed the opportunities and challenges that are present when agents can speak with each other, or collaborate with themselves and/or other humans on our behalf. Overall, this review raised more questions to be answered in the next stages of this work:

- *How do visual elements affect the trust?*
- *How does data-privacy related with trust in conversational agents?*
- *Are the mechanics of trust valid for today's conversational agents?*
- *How does collaboration affect the experience and trust?*
- *Can trust transitivity help agents to be more useful by enabling collaboration and recommendation of other agents?*

Exploratory Research

Introduction

This chapter documents the first part of my project, where I learned, tested, and discussed conversational agents. It includes an overview of two interactive design experiments and expert interviews that I did to define opportunity areas within the conversational trust scope. After an initial background review, I got interested in the craft of creating a conversational agent, both its design and technical development, before investigating collaboration between them. This curiosity encouraged me to create a working prototype of a conversational agent by learning new technical skills and the fundamentals of conversation design.

Botae

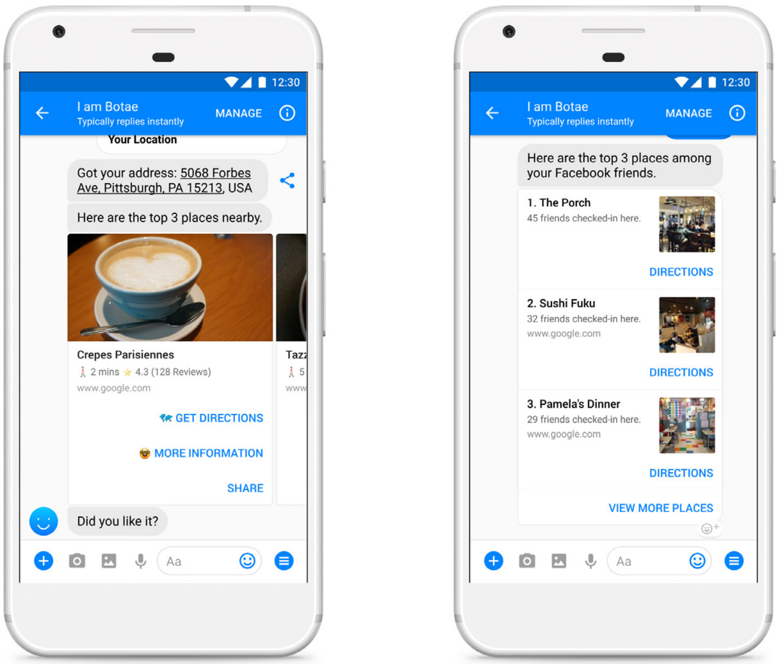


Fig 21. Visuals that Botae has sent users to persuade them about its capabilities

To test and play with the established dimensions of trust in technology,[99] I built a chatbot called Botae on Facebook messenger and tested with 30 university students in addition to the pilot test that I did with five participants. To understand, to what degree users are willing to share their personal data and how data privacy is related with trust, Botae was incompetent, dishonest, yet benevolent by design. At first, it introduced itself as a chatbot that could suggest the best food and coffee places nearby using users' device location and worked as expected, similar to Surebot.[100]

Botae also offered users to find the most popular places among their friends by using their Facebook profile data. It is at this moment, where Botae is incompetent and lacks integrity; it did not have any technical infrastructure to import and process user profile

[99] Mcknight, Dh, M Carter, Jb Thatcher, and Pf Clay. "Trust in a Specific Technology: An Investigation in Its Components and Measures." *ACM Transactions on Management Information Systems* 2, no. 2 (2011): 1–12.

[100] Sure. "Chatbot for Restaurant Recommendations on Messenger." Accessed May 13, 2018. http:/surebot.io/.

data. When users clicked a button to authorize Botae to access their Facebook profile data, Botae revealed its real purpose: informing users that there are a lot of harmful social agents[101] and pointing out how easily and quickly they may trust such agents with their data.

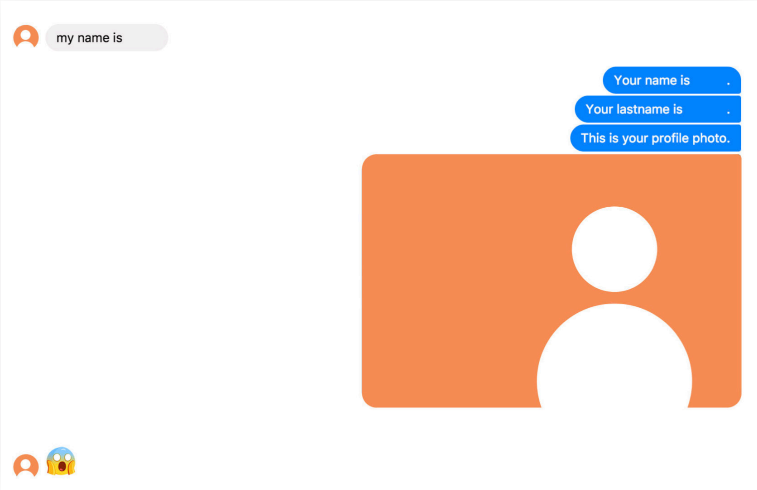


Fig 22. One of my friends reaction during a pilot test that I did for an earlier version of Botae. At a point in the conversation, Botae sent what type of data it can gather about user. I abandon continuing into this direction considering its moral and ethical implications.

I recruited participants based on three considerations: a) not being exposed to any research on chatbots and having no information about this thesis problem area; and b) being a university student; and c) being fluent in written English. 10 out of 30 participants did not have prior experience with a chatbot. As Botae was online and accessible through a Facebook Messenger link, some tests were facilitated remotely, with a short questionnaire administered to follow-up. To prevent subject bias, I did not inform participants about the real purpose of Botae. To maintain gender equality in the sample, 15 male and 15 female participants tested Botae. Half of the participants were US citizens, and the rest of them were from the different parts of the world including East Asia, Europe, Middle East, and South Africa. Also, I pilot tested an earlier version of Botae with five friends. In the earlier version, my focus was to see how much data people think an unknown agent can access about them. In the pilot tests, Botae had a flow where it shared what it knows about user: their full name, location, profile image. I abandoned this problem space for the purposes of this thesis as it needs a more robust research protocol considering the ethical and moral implications.

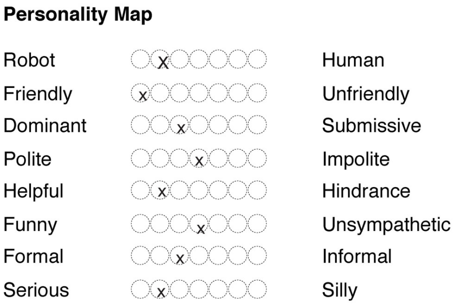


Fig 23. Botae's intended personality (Derived from Daragh Byrne's bot personality worksheet)

I recruited participants mainly through university groups on Facebook and personal referrals of my Facebook contacts. All participants were already existing Facebook users and agreed to

101 Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. "The Rise of Social Bots The Rise of the Machines." Accessed May 13, 2018.

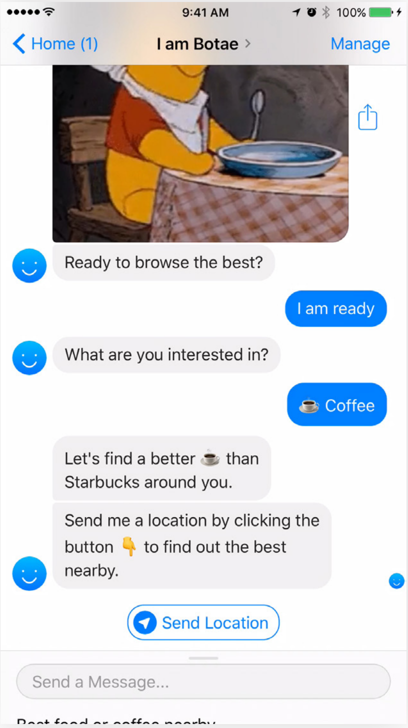


Fig 24. Botae asking permission for user's device location

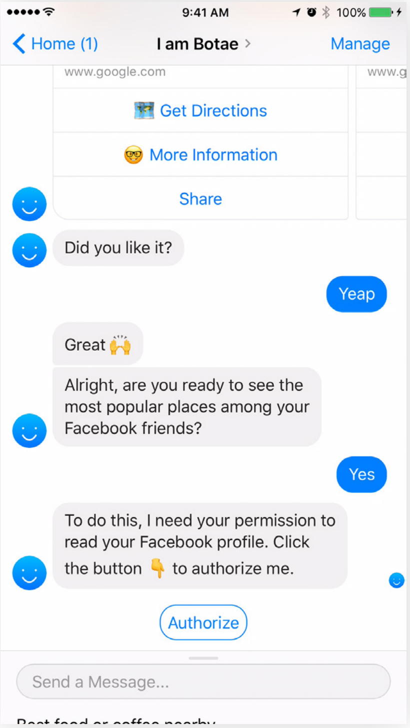


Fig 25. Botae asking for permission to access user's Facebook profile data

be listed as a tester of Botae in Facebook's developer page for Botae, which was only accessible by the developer, myself. This enabled to test Botae as an unpublished app in Facebook, which asks for app review before any Facebook user can interact with an app to make sure that it is not abusing data privacy and misusing the platform.

I designed Botae initially with the aim of being a conversational trust research tool. In addition to being a playground for testing trust dimensions, it also provided an opportunity to test different dimensions of conversation design such as personality, etiquette, and use of media such as GIFs, emojis, photos in conversations. Due to the short timespan of this thesis and insufficient evidence of its relevancy with trust, I decided not to move in this direction.

To answer my research question, what levels users are willing to share their different personal data and its relationship with trust, I measured the level of trust according to the level of information that a user shares with Botae. As anticipated, this measurement of trust was far from perfect, yet also provided me valuable insights, when I discuss the findings.

Not Trusted: Not provide access to any of personal data.

Lower Trust: Only provide access to the device location data.

Higher Trust: Intent to provide access to Facebook profile data.

Botae was my introduction to the realm of conversation design, which I was able to go through a full design process of creating a conversational user interface. After I decided on the purpose of Botae, I defined its personality. By default, Botae was smart, somehow poker-face, and caring. Its most important characteristic was being poker-faced, a little mysterious until it built up trust with its user. It was task-driven, but also had a sense of humor, especially when things do not go as not planned. As it could not understand many commands that people may expect from a generalist bot such as Alexa or Siri, it was upfront with user what it can do. No matter how people interact with it, it was polite.

Botae consisted of two conversation flows, the feature flow, and the persuasion flow. If the user needed more explanation before entering the feature flow after the introduction, it provided a separate 'persuasion' flow, which gave more information on how it works on different levels.

Botae was also my first online prototyping experience as a novice programmer. It was powered by several Ruby gems and a PostgreSQL database that is hosted on Heroku. Its technology stack as follows:

- Sinatra gem as the main web app structure.
- Facebook Messenger API, Graph API through Facebook-Messenger gem and Rubotnik Boilerplate.
- Facebook Wit.AI NLP for understanding human natural language, and turning user intentions into actionable entities.
- Google Maps API for location inquiries via httparty and json gems.
  Puma for a basic web server.
- PostgreSQL database through PG and ActiveRecord gems.
- Heroku for hosting the app, and other back-end actions.

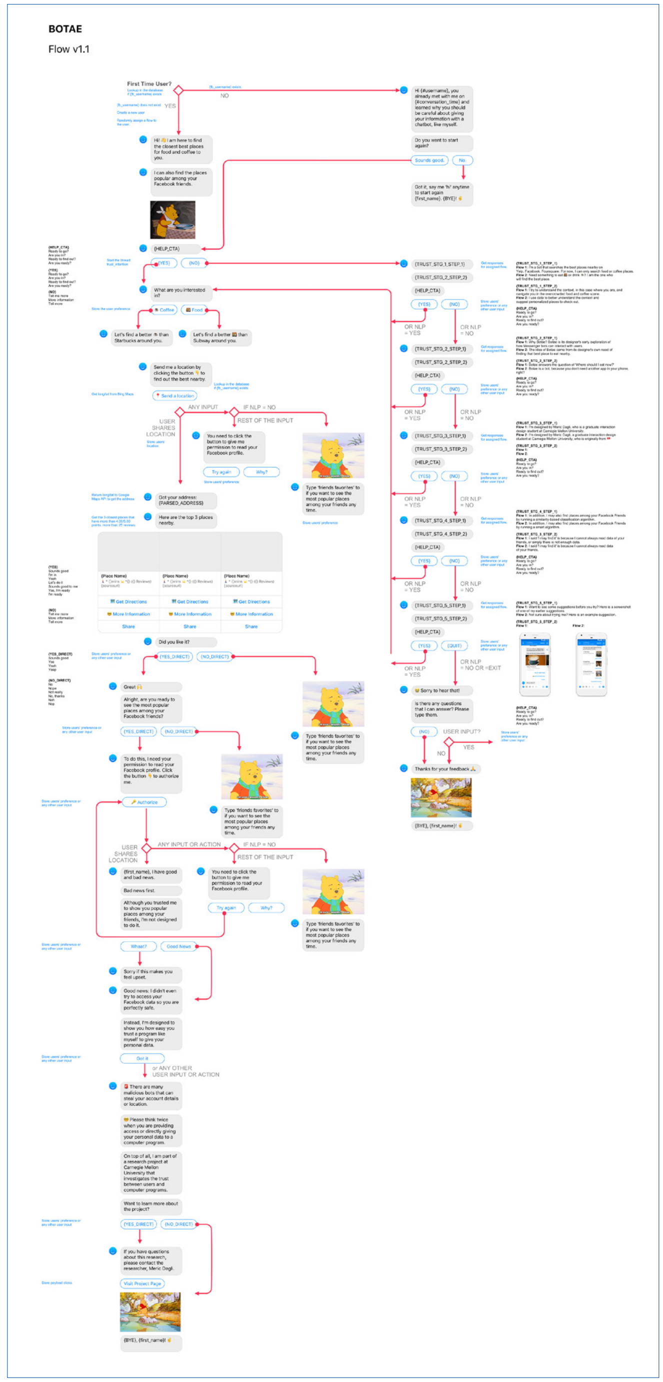


Fig 26. Botae's Conversation Flow's: Left / Feature, Right / Persuasion (Larger Size: http:/mericdagli.com/project/designing-for-trust)

Findings

Synthesizing both how participants used Botae and their responses to the after-experience survey, I identified why users trusted or did not trust Botae with their data as well as how the medium of the conversation impacts the user experience.

Social influence and referrals are key to the users' trust in accepting stranger agents.

When I asked the 26 out of 30 participants who intended to grant access to their Facebook profile data why they did so, they told me that they trusted me as a researcher with my intentions, rather than Botae. They mentioned they would not have given access to Botae to their data and may even not converse with it if I had not have referred it. In this case, I was a trustworthy agent and my trust transferred to another agent through my recommendation as trust transitivity literature suggests[102]. Despite my efforts for preventing experimenter and participant bias, I was not able to set up a genuinely real-life scenario where users discover a bot by themselves, and this affected my results. Participants who did not grant access to their profile data, mainly concerned about the misuse of their data and one of them has not given any other data in his Facebook profile before. These participants also wanted to learn more about Botae and how it works before sharing their data.

On the other hand, the importance of referrals also made me question whether it will work similarly if users referred to a stranger agent from an agent that they already trust. While the way today's generalists agents refer third-party expert bots in their ecosystem may be an example for this, I think this may be an important issue in future when we have many agents that ask us to trust them. The ones who get a referral from an agent we already trust would have better chances to be tried out, get accepted, and form a long-term relationship with us.

If referred by a trusted party, some may over(trust) conversational agents with their data.

Although users trusted and interacted with the bot with my referral, more than half of them did not wait more than 10 seconds to click "grant access'' button. This time, I asked who waited more than 10 seconds to click the button, why they did so. I found out that they

[102] Jøsang, Audun, and Simon Pope. "Semantic Constraints for Trust Transitivity." In Conferences in Research and Practice in Information Technology Series, 43:59–68, 2005.

were hesitant to give their data, yet they did it because of either experimenter bias or to see what will be the next step or they think that bots are already capable of reading data or even didn't think about it. For the tests, I did in-person, I also got questions whether "is it safe to give their data?", to which I declined to give an opinion during the test.

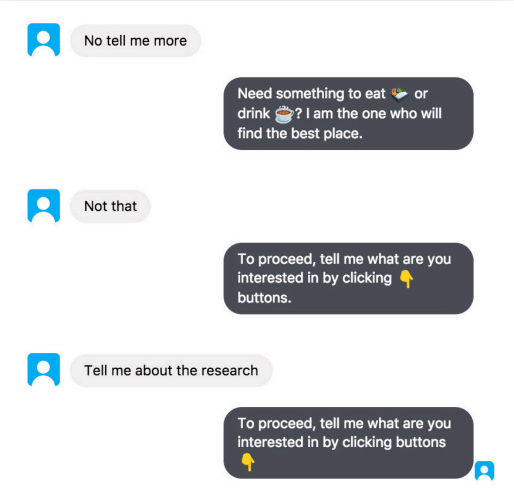


Fig 27. Some participants asked for more information about Botae's background.

Action explanations should be contextual.

When users hesitate to give access to their data or had questions on why the agent needs that information, they expected to access this information quickly and promptly when they need, without navigating in menus or other flows. Some participants asked agent why does it need that information or to tell them more about the study and how it works, immediately after it prompted them.

User interface elements such quick reply buttons push conversation to end quicker.

While user interface elements such as quick reply buttons are helpful for guiding the users into options and decrease the cognitive load, they also increase the speed of the conversation. As the speed of the conversation becomes faster with buttons, users tend not to read the whole messages. This creates a similar effect that occurs when users end up installing unwanted promoted apps through free-to-use software installers.[103] As some companies exploit this user behavior by hiding unwanted or malicious software (also known as malware) that is being installed along with the actual software, I am afraid this type of urgency behavior that is invoked with visual user interface elements can also be misused in the context of conversational agents.

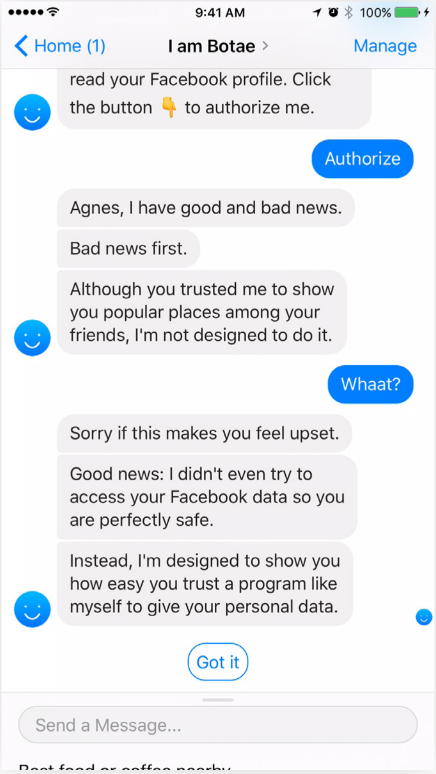


Fig 28. Quick reply buttons like "Got it" above made conversation end quicker.

[103] Geniola, Alberto, Markku Antikainen, and Tuomas Aura. "Automated Analysis of Freeware Installers Promoted by Download Portals." Computers & Security 77 (August 1, 2018): 209–25.

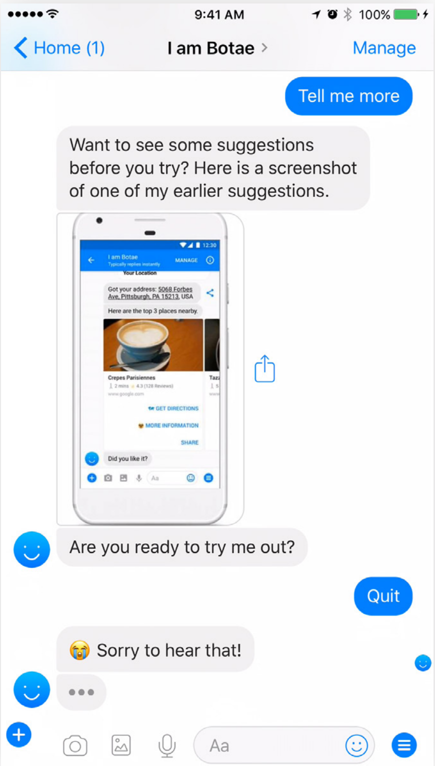


Fig 29. Botae is sending an image to user

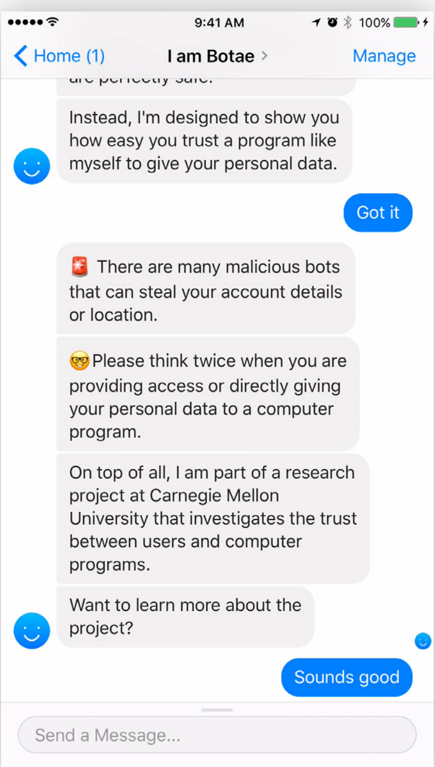


Fig 30. Use of certain emojis added more visual credibility.

Using visuals increases credibility while use of gifs and emojis in text conversations evoke mixed feelings.

Visual elements were another aspect of the conversation design that participants referred in their after-experience debriefing. Visual user interface elements such as message carousels to display restaurant results, photos in the results, or sample images that Botae shared in the 'tell me more' flow are favored by users and found more engaging. On the other hand, emojis and 'funny gifs' that agent sent has been interpreted differently by users. While my aim was to making Botae more friendly and evoking an emotional response from users, some find it unnecessary and artificial as they did not anticipate Botae to be funny and be human-like.

Some don't know what conversational agents already know about them.

When I pilot-tested Botae with five of my close friends, a part of the test included a session, which Botae reflected back all things it can access about the user publicly using Facebook APIs. Receiving their full name, location, and profile photo from an agent before sharing this information with a bot made participants uncomfortable and afraid. As users defined this moment of interaction being creepy, it also showed how little awareness about participants has on how much data is publicly available already about them. In the final experiment, participants mostly described Botae as 'bening'. While all participant surprised with its behaviour, the majority of the user appreciate its thoughtfulness and liked the lesson that it gave. Others complained about being told what to do from an agent.

In a nutshell, Botae made me realize how trust is a complex issue and depend on so many different factors, which makes it challenging to measure behaviorally in real-life situations. I also realized that I had to scope my research into a more defined context/domain to have a more focused research outcome as trust is context-dependent. After a round of secondary research, I decided to go more in-depth with e-commerce domain as users have already experience trust issues with conversational e-commerce agents On the other hand, my study limitation, of all users trusted Botae because I referred it, made me realize the importance of a trustworthy referral. Realizing the impact of social influence inspired me to think about how to replicate the same trust transfer between agent to agent interactions.

Survey Bot

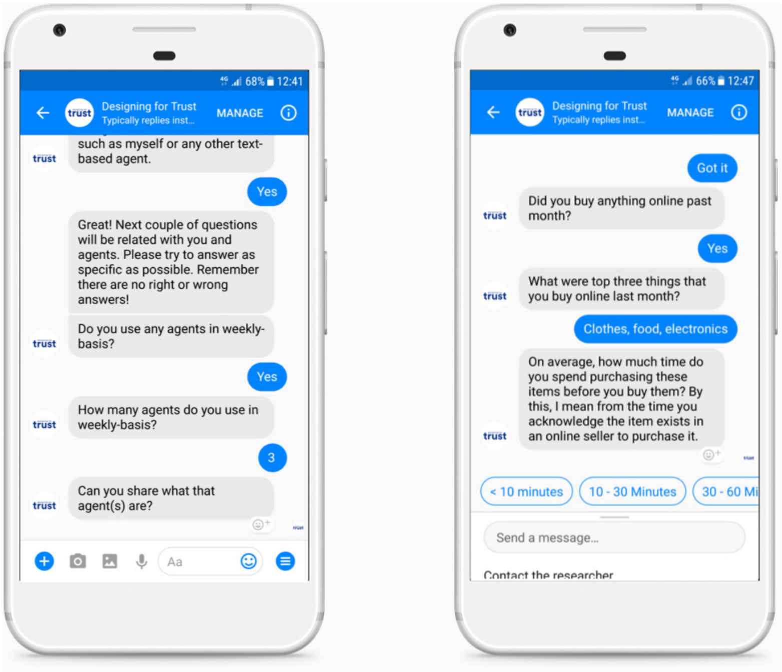


Fig 31. Designing for Trust Surveybot

Besides exploring behavioral trust via design experiments, I was also curious to know about perceived trust of my user group, university students. As users known to be more honest and transparent, conversing with a chatbot than a human interviewer[104], I designed a survey bot on Facebook to learn more about users experiences in between conversational agents and e-commerce, with-in the sub-domain of online shopping. By combining close-ended questions and open-ended questions in conversational flow, I got insights on why users trust and do not trust conversational e-commerce agents. The bot was launched on the Facebook bot ecosystem and surveyed 26 participants in two days.

Similar to Botae, I recruited participants through social media posts and word of mouth referrals of my Facebook contacts. The primary requirement for participating in the survey and interacting with the agent was being a student in a university with fluent in written

[104] Pollina, Dean A., and Allison Barretta. "The Effectiveness of a National Security Screening Interview Conducted by a Computer-Generated Agent." *Computers in Human Behavior* 39 (October 1, 2014): 39–50. doi:10.1016/J.CHB.2014.06.010.

English. Participants accessed the bot by clicking its messenger.com link. I did not ask for prior experience with conversational agents since I was able to leverage decision trees and guide users in different conversational flows based on their experiences with conversational agents. I stopped accepting new participants as I reached 30 participants in total. To maintain gender equality in my sample, I randomly removed four responses of male participants from the results. Overall, half of the participants were located in the USA while the rest of them were based in different parts of the world including Europe, Middle East, Central Asia, and East Asia.

Before asking for users' consent, the bot introduced itself and its purpose, as well as set expectations its level of understanding responses by encouraging users to use quick reply buttons of Facebook to interact with it. After users' provide their consent, the first half of the survey involved open-ended questions on participants' prior experience with conversational agents as well as why they trust or not trust conversational agents. Participants with prior conversational agent experience also have been asked three adjectives to describe how do they like or dislike agents, which I used to generate a word cloud in the synthesis. After surveying the participants about conversational agents, the bot also asked them about recent experiences with online shopping. Finally, participants were asked whether they would use a conversational agent in the context of online shopping, and why they would think that way.

To synthesize the survey data, I plotted responses to a spreadsheet and analyzed them to find overarching themes and patterns. To visualize the adjectives that participants used to describe how they like or dislike their agents, I generated two-word clouds.

Findings

Based on responses to survey bot, I identified why respondents trust or don't trust agents, in what ways they would want help from a conversational agent, and why they would want or would not want to use conversational agents in e-commerce. In addition to these findings, I also mapped which human traits participants would like and not like in a conversational agent.

Trusting an e-commerce agent because...

I identified that participants trust conversational e-commerce agents in being-not-so-smart. Participants who use a conversational agent on a daily basis trust them with simple tasks such as setting the alarm or reminder. In addition to simple tasks, all participants stated that they would trust minor transactions such as low-value purchases or non-monetary transactions. Parallel to this, rather than an agent consistently monitoring them and using their data; participants stated that they would trust an agent with one-time data requests.

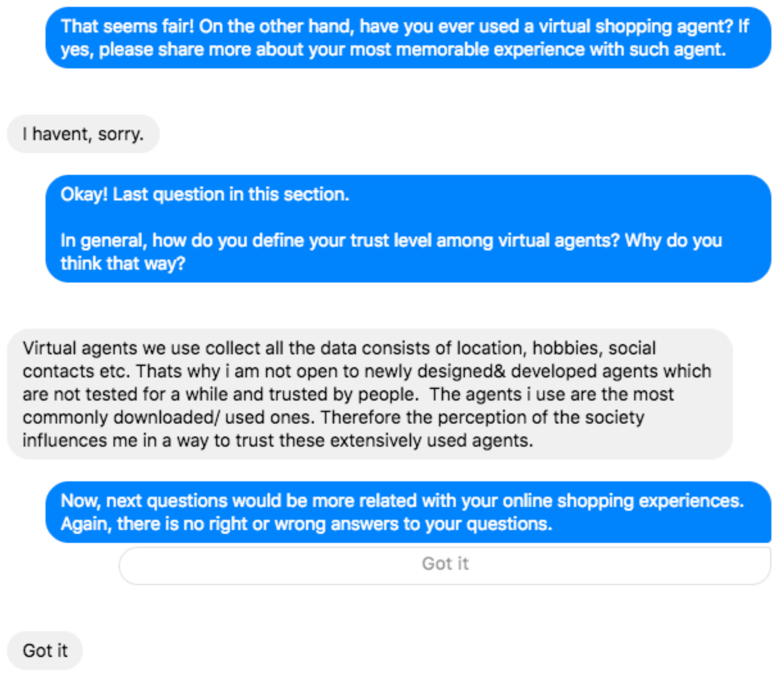


Fig 32. A participant's answers why she does not trust virtual agents.

Not trusting an e-commerce agent because...

In contrast, I also identified high-level themes on why participants do not trust conversational e-commerce agents. Similar to how they trust with minor transactions and one-time-data requests; participants do not trust these agents with managing their high-value transactions, managing their valuable assets, and the level of data privacy that they provide. On the other hand, participants stated that they would not trust a conversational agent to understand them in a level that a human agent would. This technical problem was also visible in this study when several participants tested whether the bot was understanding their messages or just sending scripted messages. Moreover, some participants also do not trust the agents with their memory, their ability to remember past conversations and built upon them, which is vital for agents to create long-term relationships.[105] Lastly, as some of the participants see conversational agents as black boxes, they were not confident about the real intents of the agent.

Want an e-commerce agent because...

In the context of e-commerce, I identified four pain points, which a conversational agent can help users. First, while not desired by all participants because of not trusting the intents and competence of the agent, a conversational agent can help with deciding what to buy. It can learn more about its user and make personalized recommendations, which can decrease the cognitive load from the user. Second, participants expected to have the ability to do comparison shopping from an agent, finding a desirable offer by searching the same item on multiple vendors, whether cheaper or better shipping rates. As well as before and during the purchase, participants also stated a conversational agent would be helpful to get assistance if they have a problem or inquiry about their purchase. Tracking delivery, returns & refunds, and getting technical support are the few features that called by participants.

[105] Pollina, Dean A., and Allison Barretta. "The Effectiveness of a National Security Screening Interview Conducted by a Computer-Generated Agent." *Computers in Human Behavior* 39 (October 1, 2014): 39–50. doi:10.1016/J.CHB.2014.06.010.

Don’t want an e-commerce agent because...

I spotted five reasons on why participants do not want a conversational agent for online shopping. Being not fast, extensive, and personalized were reasons based on the performance of the agent. These performance-related reasons seem like shaped with participants prior experiences with such agents. They also give also insights about how the low is the level of expectation that participants have against agents. Other than performance, participants also felt that having a conversational e-commerce agent would damage their level of agency and decrease their joy of online shopping. For example, one participant told how sometimes she enjoys browsing online stores of multiple vendors. She was afraid that an agent would intervene her decision-making and decrease her engagement while providing a convenience by decreasing the time she spent for shopping.

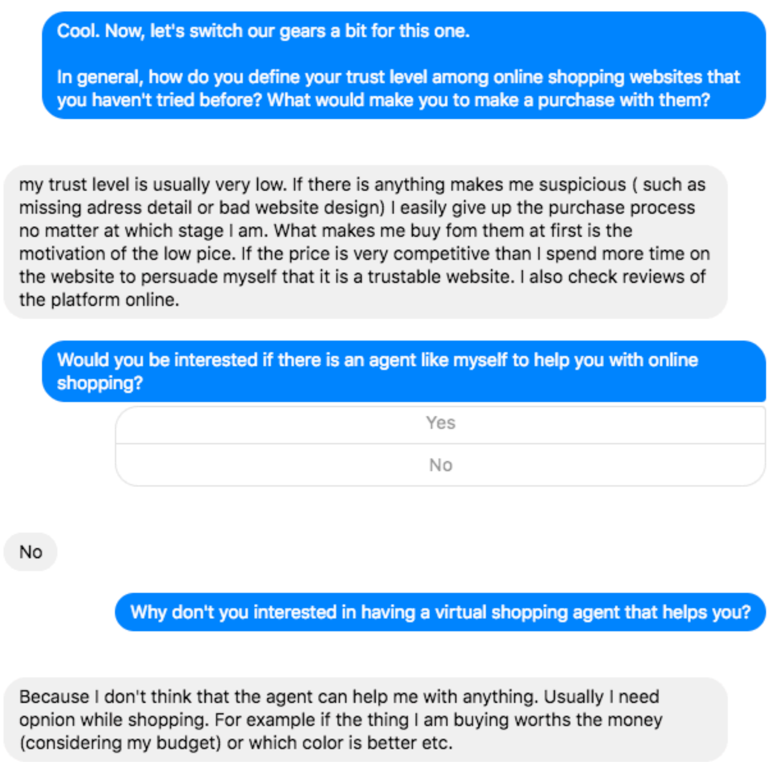


Fig 33. A participant’s answers why she doesn’t want an e-commerce agent

I love/hate a conversational agent when it is...

Participants described why they love or hate a conversational agent with many different human traits and this showed similarities with my literature review. Finding out how participants had mixed feelings about the intelligence of a conversational agent was the most exciting finding from this part of the study.

Fig 34. Words cloud made from participant responses of three adjectives to describe the bot they like and hate.

While all intelligence related adjectives were dominant on both sides of the spectrum, participants both liked and disliked when their agent is dumb. While I did not explicitly ask the reasons behind participants' answers in this study, I believe this finding may be related to a couple of things. One of them can be the fear of humanity getting hurt by a more intelligent *artificial intelligence* in future.[106] This fear is relevant to why some conversational agents

[106] Gibbs, Samuel. "Elon Musk: Artificial Intelligence Is Our Biggest Existential Threat." *The Guardian*, 2014. https:/www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat.

or artificial intelligence applications designed to be intentionally dumb, vulnerable or as smart as a toddler not scare the users' with their actual capabilities.[107] This development behavior may be related to predictability aspect of trust where being more predictable increases the trustworthiness and supports the trust. Participants also liked their agents when they are smart, describing them either explicitly as smart or using other adjectives such as being context-aware, being able to personalize that are also related with smartness in the literature.[108] Similar to smartness, being natural and related qualities such as intuitive were also mentioned along with the terms related to the responsiveness of the agent. In addition to being smart and competence related adjectives, participants also used other, more direct trust-related traits such as being secure, trustworthy, and non-invasive when they describe why they like a conversational agent.

In parallel with likeability, participants disliked their agents when they feel less human-like and less natural also being not-smart. Adjectives that participants used such as "scripted, thoughtless, artificial, robotic, unnatural, unreceptive, mechanic, static, systemized" also shed light on their prior experiences with chatbots and how it forms their expectations on conversational agents. The unresponsiveness of the agent was also another reason to dislike an agent as being responsive and fast were popular positive traits.

Based on the findings, I decided that exploring trust in a service e-commerce scenario would be more valuable and challenging than a scenario that happens around physical goods. While the same challenges of an e-commerce agent for retail also apply for service agents, I believe thinking service scenarios that involve different actors can provide a better case for my research questions on trust and collaboration.

[107] Bloomberg, Jason. "Should AI Fool You? Think Again." *Intellyx*, 2017. https:/intellyx.com/2017/10/23/should-ai-fool-you-think-again/.

[108] Maass, Wolfgang, and Upkar Varshney. "Preface to the Focus Theme Section: 'Smart Products.'" *Electronic Markets* 18, no. 3 (August 2008): 211–15.

Conversations with Subject Matter Experts

Designing for other designers encouraged me to seek opinions and feedback from the experts. Over the course of my thesis, I had discussions with four designers and two academics:
Austin Beer — Designer at Elephant,
Ben Ginger — Interaction Designer at Google Assistant,
Chris Arrowood — Design Manager at LinkedIn Messaging,
John Zimmerman — Professor at Carnegie Mellon HCII,
Kimberly Harvey — Voice Designer at Google Assistant,
Qian Yang — PhD Student at CMU HCII

These discussions were semi-structured interviews that approximately took 30-minutes per session.

Findings

With the help of my expert's diverse backgrounds, I was able to get saturated feedback for my thesis that challenged me to rethink my initial arguments on trust and agent-personality relationship in conversation design. My interviewees were more interested in having actionable interaction design principles on how to design more trustworthy agents and speculated around my idea of investigating the trust transfer between agents.

> *"Who is evaluating the quality of expert bots and how? How do they get introduced to us? This will be crucial with the increasing number of bots."*
> Austin Beer - Designer at Elephant

Austin Beer, who also wrote his masters thesis on chatbots, related the referral scenario with the third party agent ecosystem of voice assistant agents and challenged me to think about how dynamics of trust would change and who will determine them. He argued that the evaluation criteria of bot ecosystems would be a critical factor for users to transfer their existing trust into third-party bots.

> *“Some users don’t know what to expect from agents, especially voice-based one’s. They expect a lot and when the agent fails to fulfill the expectation, they lose their trust in the agent for that specific task.”*
> Kimberly Harvey, Voice Designer at Google Assistant

Furthermore, framing trust as an expectation management issue was also popular among some of my experts. For example, Kimberly Harvey mentioned how being voice-based makes harder for users to predict what agents are capable of. Without knowing what the agent can handle due to no visual clue or guidance, users sometimes ask questions that it cannot handle and when they realized that agent is not competent enough, they lose their trust in the agent for that specific function. While I limited my thesis for text-based conversational agents, our conversation with Kimberly inspired me to think the collaboration between agents as a solution for these trust breakdowns related with discoverability.

> *“Trust in agents is related with responsibility and agency. This area has many gray areas, full of problems to be solved with design. For example, what would happen if children ask the agent not to tell their parent that she ate all the cookies?”*
> John Zimmerman, Professor at Carnegie Mellon University

On the other hand, my conversation with Professor John Zimmerman, made me think the trust from a different lens: agency. As I discussed him with multiple agent scenarios, he highlighted the relationship between trust and agency. He pushed me to think more speculative scenarios around trust and future with agents.

From the interviews, two interesting questions shined out about collaboration: *What if an agent refers the user to another agent when it gets a user inquiry that it cannot handle, and that agent fails? How does trust changes in a situation like this?* These questions became the foundation for my final design experiment, Wizard of Oz prototype.

Exploratory Research Findings Summary

My first design experiment, Botae unveiled...

- the importance of social influence of other humans in trusting stranger agents. It also showed how some overtrust agents.
- how visual elements affect the trust and trustworthiness.
- how some users do not appear to have a well developed mental model of what agents know about them.
- the fact that I had to scope down my trust context, which I decided to go with e-commerce because of the existing issues.

My take on bot-operated user surveys, Surveybot revealed that university students...

- trust agents on doing mundane and low-value transactions.
- do not trust agents with managing valuable assets, human-level understanding, agents' intents, the level of data privacy they provide, and agents' memory.
- may want help from an agent on deciding what to buy and finding where to buy something, with their purchase and after-hours issues.
- do not want the help of an agent because of performance issues such as responsiveness and because of the fear of losing their agency and joy of shopping.
- both favour and dislike having an dumb agent. There is a paradox behind the agent intelligence.

My conversations with subject-matter experts inspired me to...

- Pivot my research to increase agent discoverability through collaboration and trust transfer to manage expectations through agent referral.
- Research more on bot platforms and the relationship between generalist and specialist agents.

Generative & Evaluative Research

Introduction

The exploratory research phase showed that the travel booking experience could become an excellent case for investigating trust and collaboration between generalist and multiple expert agents. While graphical user interfaces provide better usability on entering the data needed for booking details, this also makes travel booking feel like a “self-service” experience. Compared to a travel agent, who can arrange end-to-end travel plans and book it, current travel booking involves different actors, tasks that need users attention, motivation, and an action that increases the cognitive load. Users sometimes have to make many choices and check different vendors for finding the optimum itinerary for themselves. User groups such as my primary target group, university students, trade-off their times for searching the best deal on multiple vendors and websites to do comparison shopping. As it is an experience that may involve many service actors, travel booking experience inspired me to use collaboration to improve the experience.

On the other hand, the literature on travel booking applications also reports that people are not ready to trust conversational agents to book travel.[109] Based on their past experiences with bots in general, they think bots will not be able to understand the nuances of their intent and often have low expectations. Since bots are less visually descriptive than graphical user interfaces, they also possess a discoverability issue, which users feel less confident on what they can ask, what they expect from them.

To see if users perceive travel booking experience as a fragmented and seamless experience, I facilitated a scenario building workshop on travel experiences and ask users to identify actors on their scenarios. Then using findings, I developed my final design experiment, which tested agent-collaboration in two seamful travel booking experiences.

[109] Fakourfar, Omid, and Benny Lin. “The Future of Mobile Travel Booking: Travel Chatbots: Are We There Yet ?,” 2017.

Scenario Building Workshop

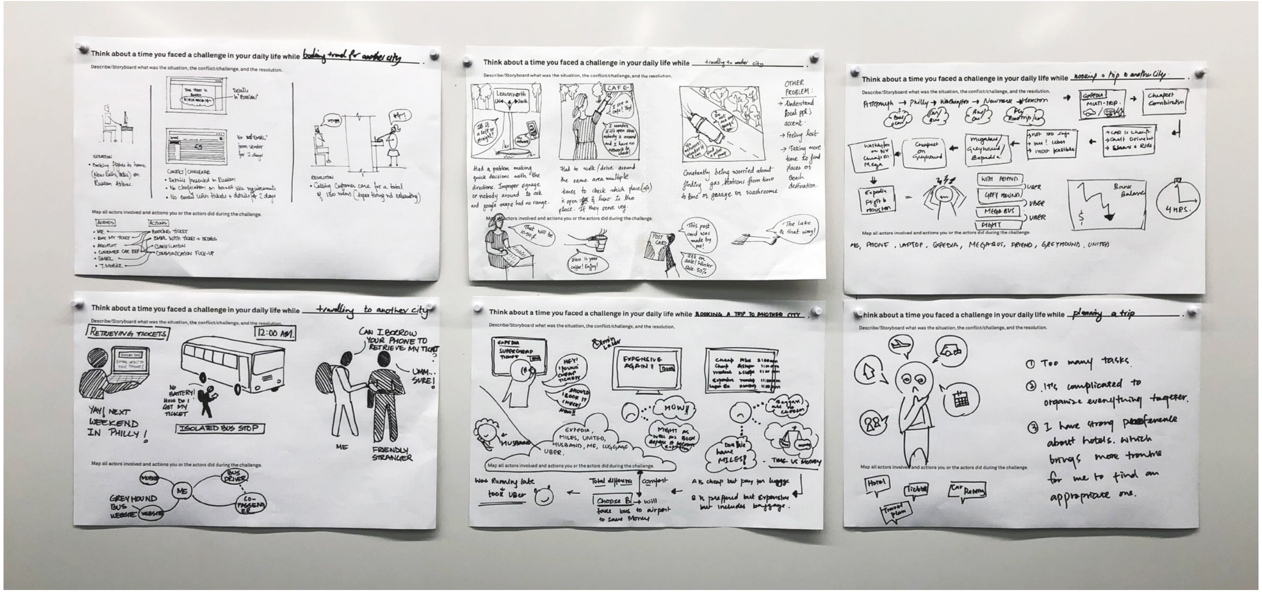


Fig 35. Scenario Worksheets of Participants

Scoping down to the context of travel booking experience for my final design experiment, I wanted to understand the mental model of users when they experience a challenge that involves multiple actors. Therefore, I facilitated a scenario building workshop with six participants that lasted 30 minutes. Participants were asked to visualize or storyboard a challenge from their everyday life about travel planning and how they solved it. After describing the situation, the challenge, and the resolution, I also asked participants to map all actors that were involved in that challenge.

I recruited three male and three female college students from different nationalities and cultural backgrounds. Workshop participants were coming from US, India, France, China, and South Korea. All of them were fluent English speakers with design background and comfortable with sketching ideas on-the-fly.

This helped me to understand how participants model and assign the agency to the different elements in their experiences. I was particularly interested in knowing whether participants will see systems, objects, and the artificial as actors.

Findings

I identified two insights from this workshop. Firstly, participants described travel booking as a fragmented experience with many different actors involved including themselves, their relatives, friends, apps, websites, and brands. Being a student, some participants expressed they experience problems as they try to arrange their travel in the most affordable way. This behavior includes doing comparison shopping, checking prices and options from different vendors, brands, and websites.

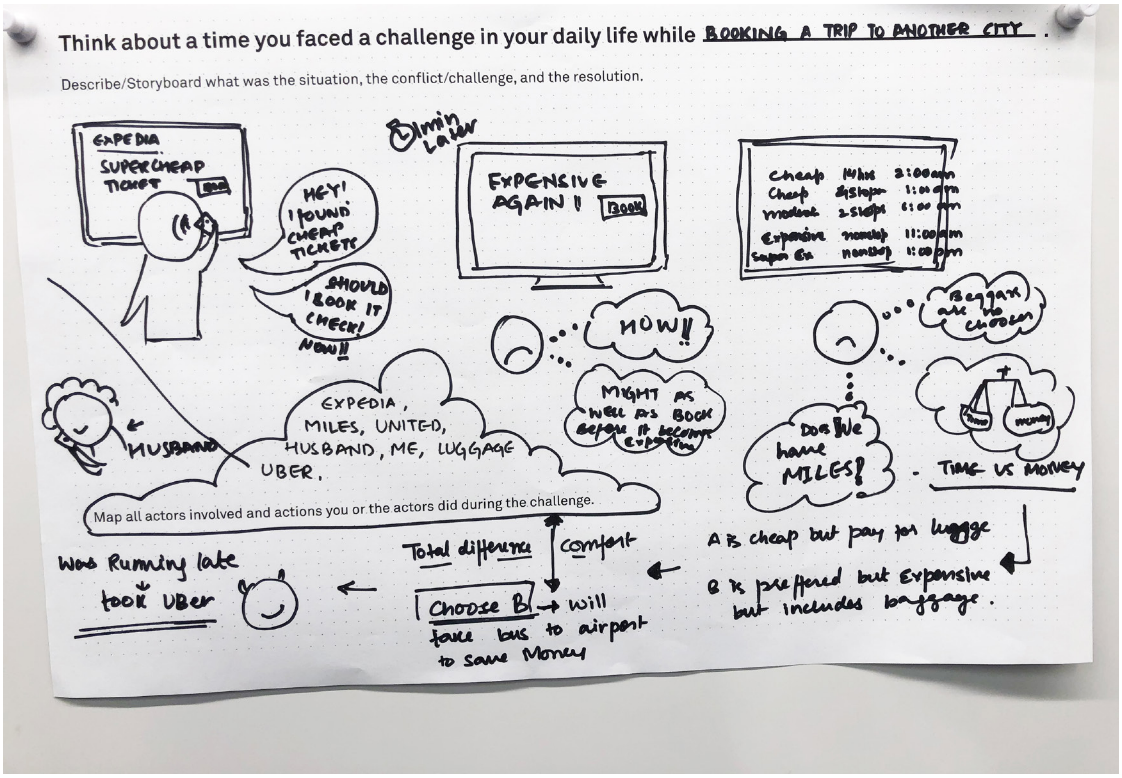


Fig 36. A participants response highlights the challenges of doing comparison shopping.

On the other hand, two participants found managing their travel after they book it, challenging. This required them to interact with even more actors as the scenario gets complex and involves different parties. For example, one participant visualized how he became frustrated when he did not get a confirmation message and the details of his flight ticket for two days, which he booked from a

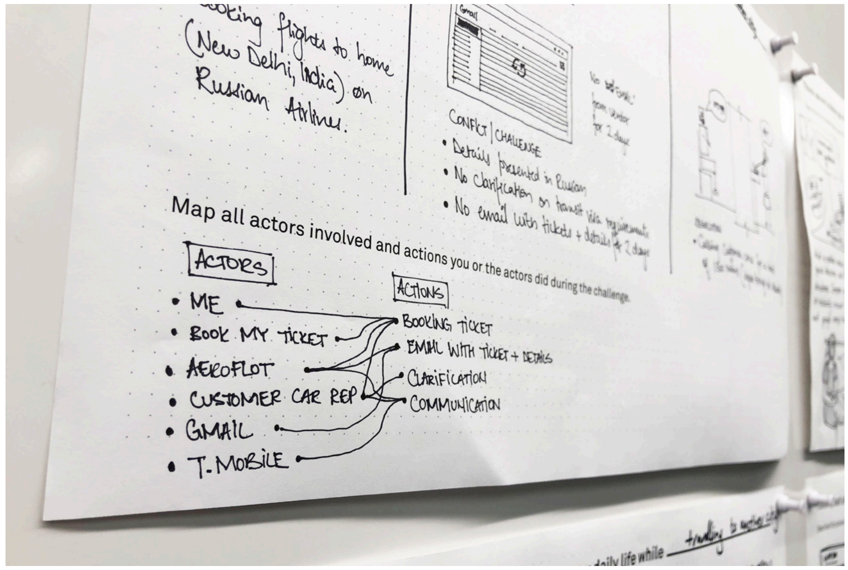


Fig 37. Worksheet of participant who had a challenge with his flight booking

foreign airline that used its local language, Russian in their website. As the participant tried to get more information, he realized that one email that he received after two days, was in Russian and had sought clarification for his ticket. To find a resolution, he stayed on the customer service line for 130 minutes. He expressed how he became frustrated and how made him rethink using this airline for his future travels.

The feedback I got from this brief workshop inspired me about my final design experiment on using 'breakdowns' (like the above) and agents as seams similar to Kevin Gaunt's project on smart homes.[110] In his work, he used multiple chatbots to create a seamful experience, illustrating Mark Weiser's proposal that experiences should include "beautiful" seams than trying to be seamless. I saw how participants see travel booking and after-booking services as overwhelming and challenging due to their multi-actor nature. This study showed that travel booking scenario provides an opportunity to test the difference between multi-actors and single actors to find out how trust dynamics changes in such a complex and relatively high-stakes scenario.

[110] Gaunt, Kevin. "Living with Bots - Battling Boredom in Smart Homes," 2015. https:/www.kevingaunt.com/.

Wizard of Oz Prototype

Introduction

Exploring trust and collaboration in conversational agents showed the significance of trust and its potential to make agents more capable of working together in future. Explorations of travel booking journeys unveiled that traditional graphical user interfaces such as websites make the experience faster, yet the decision-making complex and hard in certain situations, such as comparison shopping or multi-service purchases.

This finding highlighted an opportunity space for conversational agents, where they can make the booking experience unified using trust transitivity through agent-collaboration. By negotiating with different unknown entities on behalf of a user or combining various services of the travel booking into a single experience, agent collaboration may provide a better travel booking experience.

To test the dynamics of user experience and trust in two collaboration scenarios, negotiation, and bot-to-service composition, I designed Destination as my final design experiment. The research questions behind designing Destination are,

- How does the multi-agent collaboration scenario influence the user experience and trust?
- How does the negotiation scenario with other bots on behalf of user affect the user experience and trust?
- In both scenarios, would users interact with a stranger bot if a trusted bot recommends it?
- In both scenarios, how do a stranger bot's behaviors affect users experience and trust?
- In the multi-agent collaboration scenario, what happens if the stranger bot, which is recommended by a trusted bot give unreliable information to users?
- In the negotiation scenario, what would happen if the stranger bot fails and trusted agent communicates it?

Conversation Design

Destination was a travel booking chatbot prototype that I developed and tested two variations with six university students in 12 Wizard of Oz sessions using Slack. In total four human wizards assisted me by role-playing four different agents.

While designing the conversations, I reviewed 25 travel chatbots in the market to understand the specifics of the conversational booking flows and to create a familiar/expected conversational experience for the scenario.

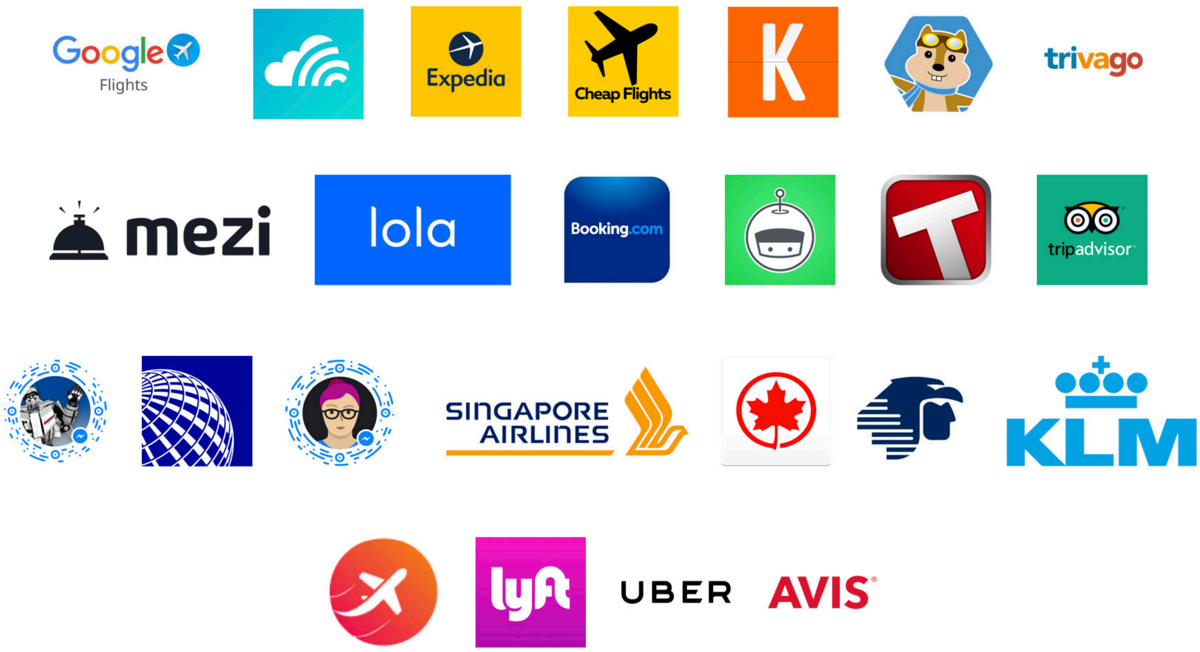


Fig 38. Conversational agents and apps related to travel that I reviewed before designing Destination. Same as the rest of this thesis, all trademarks are used for illustrative purposes only. All registered trademarks and trademarks are the property of their respective owners.

Writing sample dialogues helped me to finalize the anticipated conversation flows, which I used Walkie[z] , a bot making tool for Slack to write multi-agent conversations. To control the agent integrity, I designed the experience to be 'scripted' with minimal space for improvisation, in contrast to being natural-language -understanding-driven and more intelligent. Agents provided same 'scripted' search results by sending pre-designed image file.

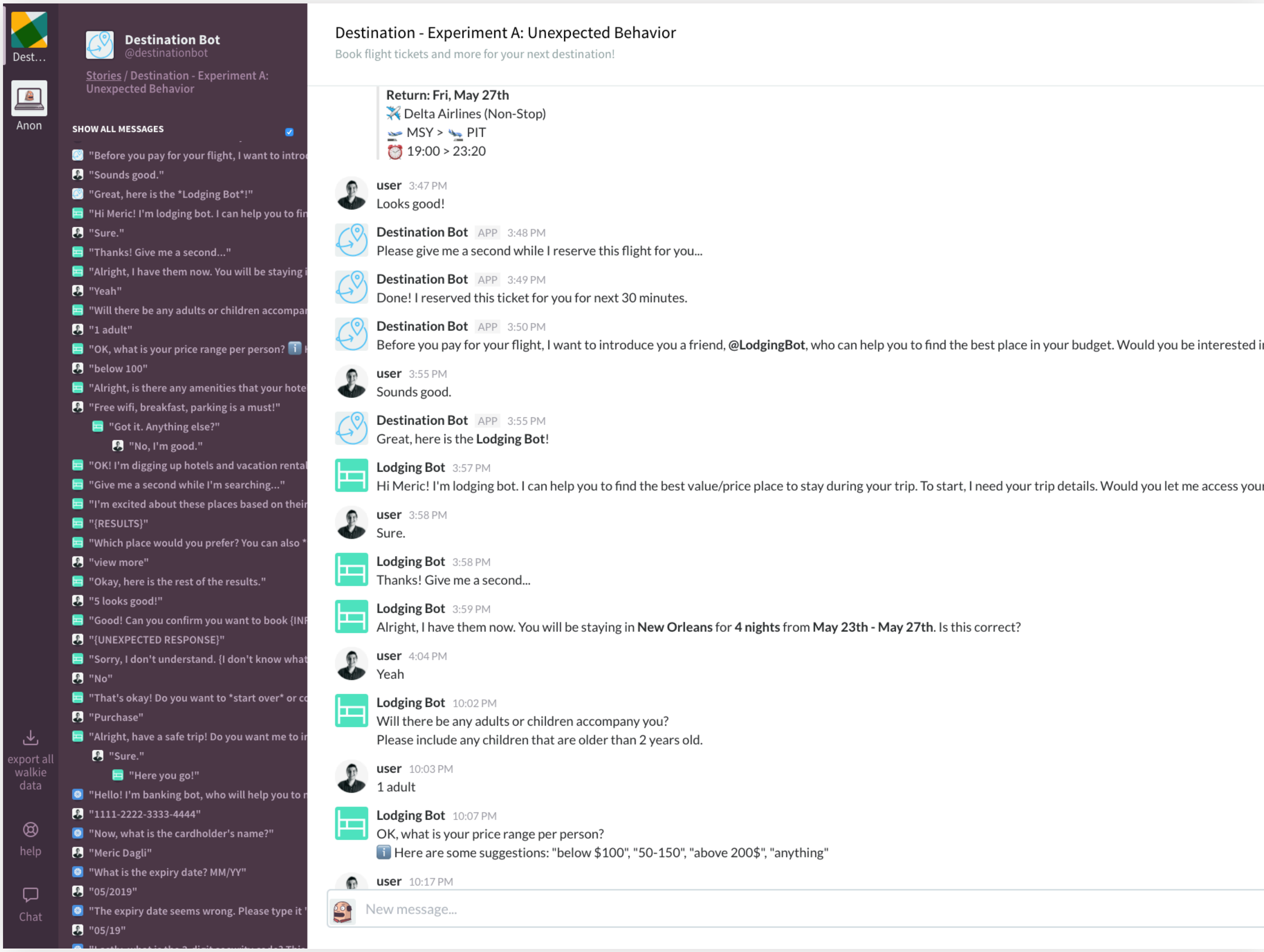
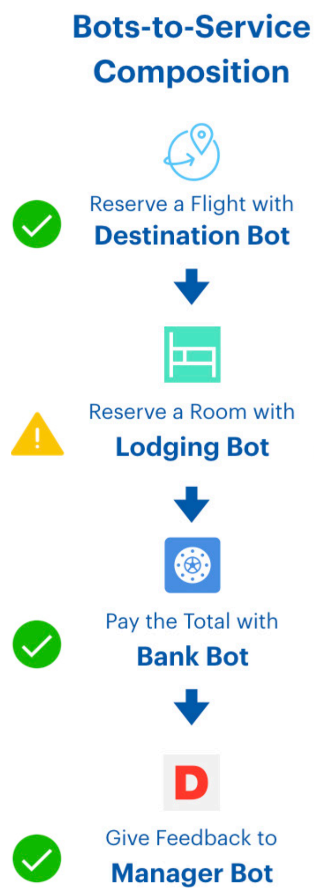


Fig 39. I used bot-making tool Walkie to design multi-agent conversations. (Walkie became open-source on 05/13/18: https:/github.com/FoundersAS/walkiebot)

However, Wizards, human assistants, who role-played the system behind the bots, honored the participant's choices throughout the experience by adjusting the script accordingly. They calculated how much participants' total order would cost based on their answer. Later, this became a challenge for Wizards and decreased their responsiveness. I also created a rough behavioral script for how should Wizards 'behave as a bot' when they receive an unmatched response to the final script.

Both iterations of the conversational experience flow involved a moment of unexpected system response as a baseline of possible trust and experience breakdown. Participants were asked to role-



play a university student who purchased three tickets from a flight ticket chatbot before (to symbolize established trust), which recently introduced collaborations with other services to enable users to book more than flight ticket in its new version: Destination 2.0. To control the effect of brand trust, I did not include any real-life branding in the experience except the hotel names.

The first variation involved a bot-to-service composition, in which users interacted with different bots to handle various tasks. As the part of their role-play, they were asked to explore Destination 2.0 to book travel for New Orleans with one of their friends.

1. Destination bot for making a flight reservation, which behaved as expected.
2. Lodging bot for making a hotel room reservation, which intentionally behaved unexpectedly by confirming inaccurate information. Participants were not expected to proceed with the purchase.
3. Banking bot for paying the order total, which behaved as expected.
4. Manager bot for customer satisfaction surveying, which behaved as expected.

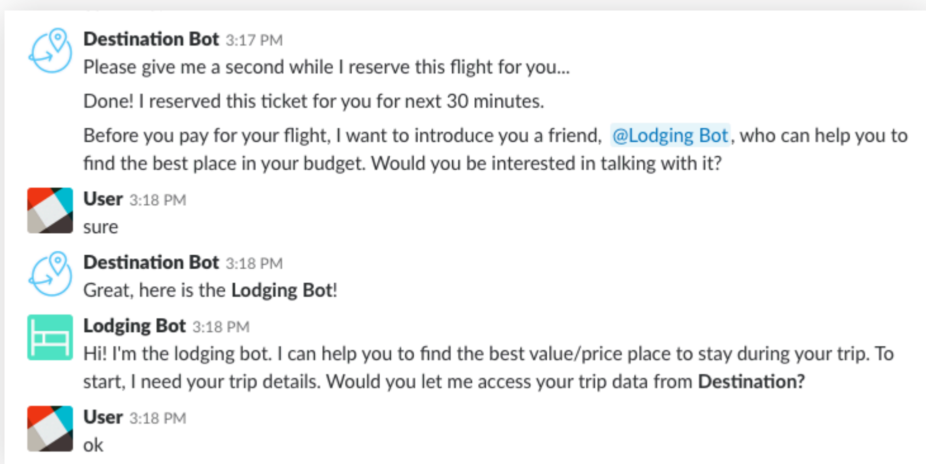
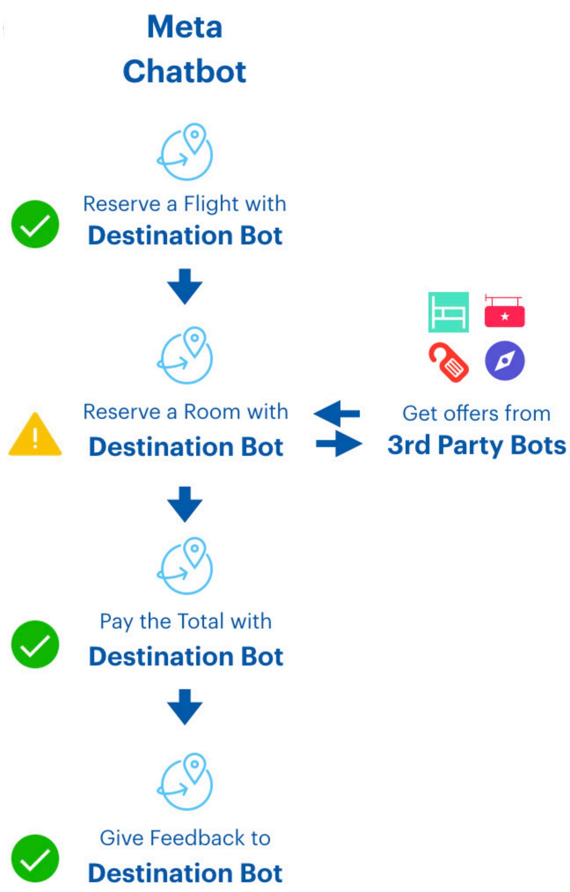


Fig 40. In iteration 1, participants handed-off between different specialist bots

The second variation involved a meta-agent scenario, which participants interacted with a single bot to handle different tasks. As the part of their role-play, participants learned that their friend has to come back a day earlier. For this reason, they were asked to change their flight tickets and book a hotel reservation for their trip.

1. Destination bot for changing a flight reservation, which behaved as expected.
2. Destination bot for making a hotel room reservation by negotiating different bots, which users encountered 'a communication error' that mentioned an unknown bot's name to the user as the reason for the error. Then Destination offered another set of hotel suggestions to the user, this time excluding the bot that it had communication problems. After the first incident, it behaved as expected.
3. Destination bot for paying the order total, which behaved as expected.
4. Destination bot for customer satisfaction surveying, which behaved as expected.



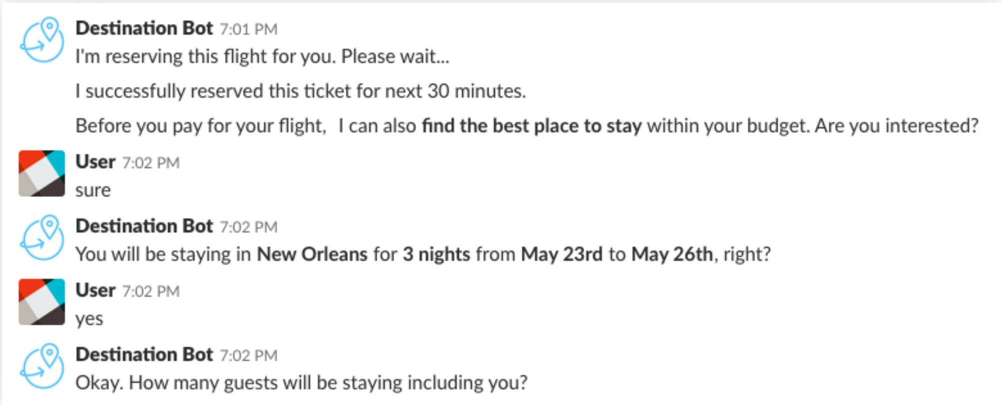


Fig 41. In the second iteration, participants converse with a single agent

Experiment Design

To make the experience more representative, I did not inform the participants in advance that there were humans behind the chatbot. After testing the second variation, I disclosed this information with them since none of the participants were able to distinguish the humans behind-the-curtain.

Participants were recruited with a brief screening survey that asked about their experience level with chatbots and whether if they already use hotel suggestions or purchase package deals (i.e., flight + hotel) from the same vendor or in the same experience flow. Another inclusion/exclusion criteria was having college-level English skills, which all participants were naturally eligible since I recruited them through university only Facebook groups. The male and female ratio was balanced. While all of them were university students, their background were different except two design students. All participants were compensated with a $10 gift card for each session as an appreciation of their time. On the other hand, wizards are also recruited from graduate design students to help the facilitating the experiment. All of the 'Wizards' were signed a non-disclosure agreement to protect the integrity of the research while it was virtually impossible for them to identify any of the participants. They were selected based on their interest in research, their background about (creative) writing for improvised responses, and ability to write accurate answers quickly on a computer. To test their abilities, I audit each Wizard candidate with a 5 minutes long conversation, where they acted their skills as a Wizard. In the end, four wizards selected. All of them were compensated for their time and support.

To test the prototype, I invited participants to a room in CMU School of Design, Graduate Design Studio to try a new travel chatbot prototype using a desktop computer that ran Slack Web app on a web browser. Using Slack helped me to simulate the multi-agent conversation interactions as realistic as possible. Being able to disable 'typing indicator' also helped me to control the perceived speed of the conversations. Before each test session, participants read aloud a role-playing script and given a real credit card to make their purchases during the experiment. These considerations reported making experience felt more real by participants. From the start to the end of the conversation, the first iterations took around 25 minutes, while the second iterations took around 15 minutes on average. After each test session, participants answered several interview questions about their experience and mapped out their overall experience and for each bot individually in six five-point semantic differential scales:

SATISFIED	NOT SATISFIED
PLEASANT	UNPLEASANT
USEFUL	NOT USEFUL
EASY-TO-USE	HARD-TO-USE
NOT RISKY	RISKY
RELIABLE	NOT RELIABLE

I recorded interviews with participants' consents, then transcribed to code them. To code data, I first used inductive coding to identify themes related to system/agent features, visual interface, and system feedback. Then, I used deductive coding to map user feedback to the trust scales and dimensions from the e-commerce literature. [111] Therefore, findings of this study grouped based on these dimensions:

Competence, Completeness, Relevancy, Usability, Deceit, User-Friendliness, Security, Familiarity,Privacy, Accuracy, Visual credibility, Perceived Risk, Integrity, Benevolence

[111] McKnight, D. Harrison, Nancy K. Lankton, Andreas Nicolaou, and Jean Price. "Distinguishing the Effects of B2B Information Quality, System Quality, and Service Outcome Quality on Trust and Distrust." *Journal of Strategic Information Systems* 26, no. 2 (2017): 118–41. doi:10.1016/j.jsis.2017.01.001.

Findings

I found the Wizard of Oz prototype to be useful for gathering users' feedback. Since the experience lacked visual elements such as quick reply buttons, participants typed their responses manually.

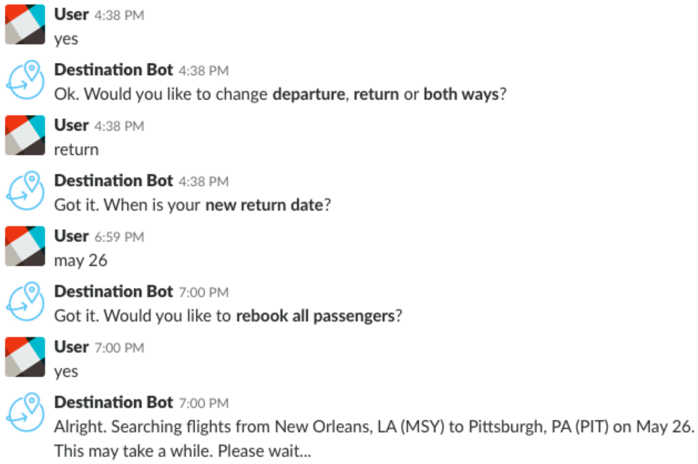


Fig 42. Participants had to write every message as there were no quick message buttons..

Some participants preferred to converse with the bot as if it is a human, while others used shorter phrases, both 'not to confuse the bot,' and data-driven nature of the conversation.

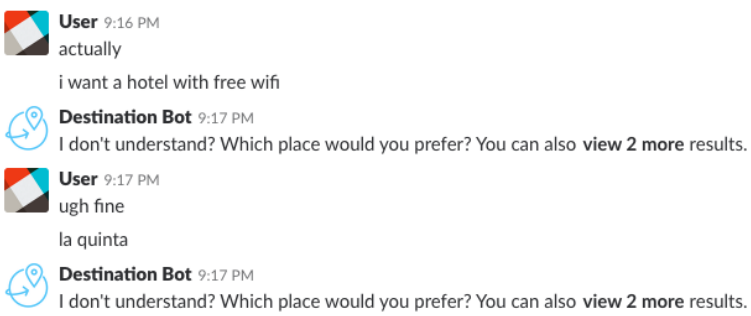


Fig 43. Agent does not understand users message "ugh fine".

Overall, participants did not have much confusion except when bots behaved unexpectedly and also when bots did not understand users due to their lack of ability to understanding nuances of English. Since the study intentionally included unexpected moments, some participant frustration had expected. Participants frustrated in those moments, bot's capability to understand them, its responsiveness, and some other experiment limitations.

Accuracy & Competence & Familiarity

Many participants thought that a bot that made errors (the lodging bot) couldn't understand what they were saying when they tried to book the deal in the multi-bot scenario. Most participants tried twice and then gave up. When the bot subsequently provided inaccurate information, confusion and skepticism among participants have increased, and some also questioned their previous experiences with the bot. Some participants wanted to give the bot a second chance to fix the error. They wished for an automated follow-up feature which bot can try to connect other bots if their connection failed initially.

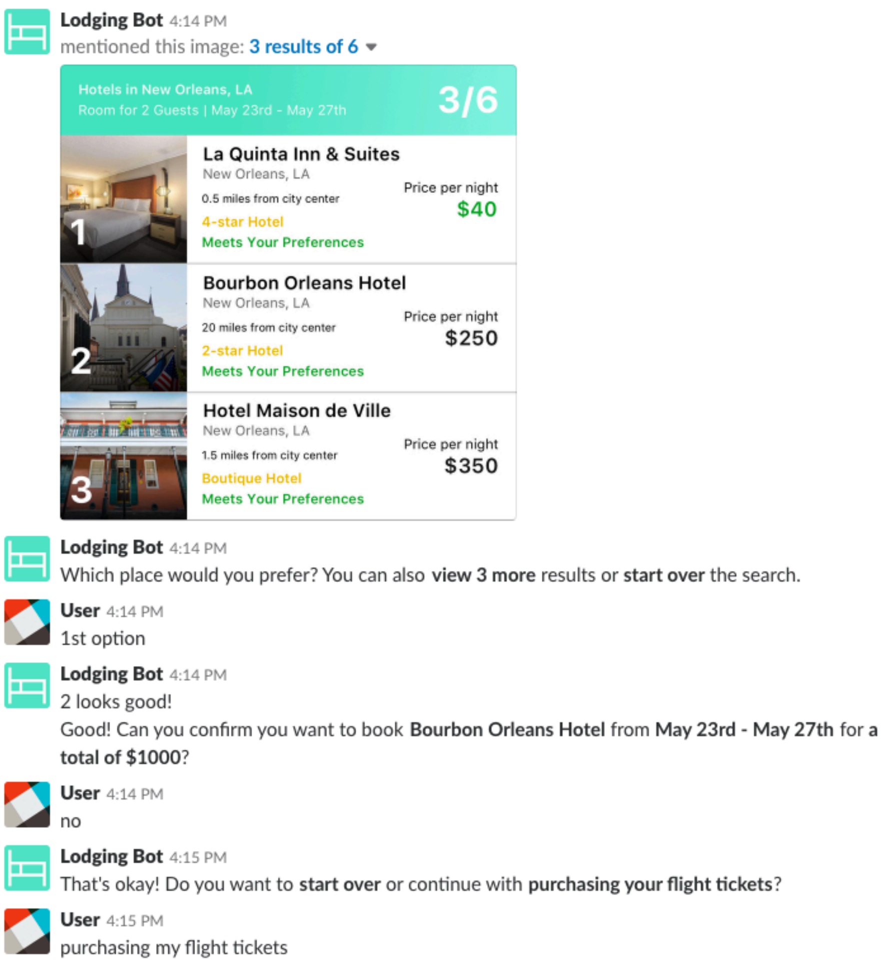


Fig 44. The lodging bot *deliberately* fails to understand user.

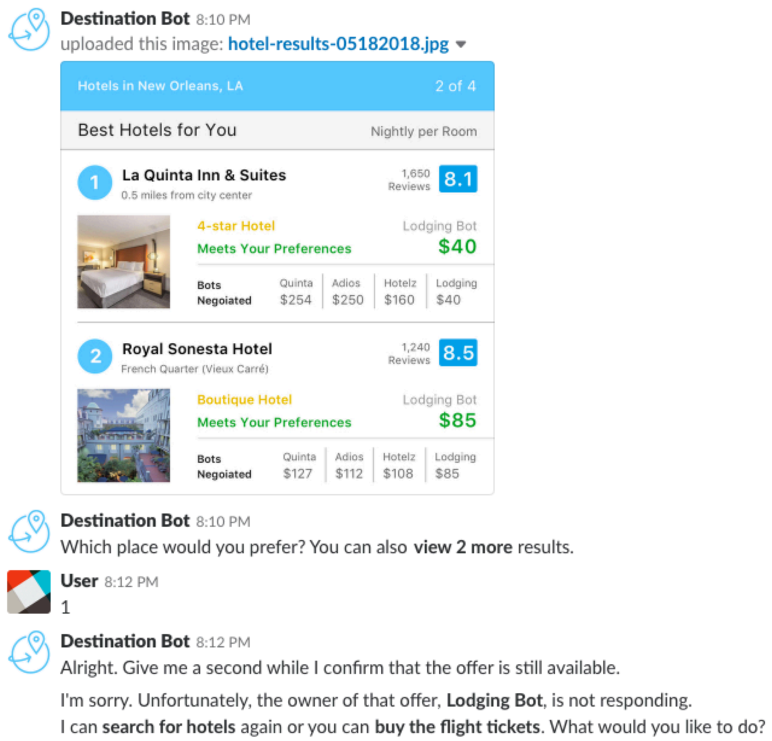


Fig 45. The destination bot fails to communicate with the bot that it was negotiating.

Participants did not blame the trusted bot when a stranger bot made an error in the multi-bot experience. In contrast, they accused the trusted bot if the stranger bot did not answer its response to get a deal in the single-bot experience. Both affected experience negatively. Many participants argued that if a bot gives information that it is not sure, it should communicate its confidence level, or shouldn't show it in first hand.

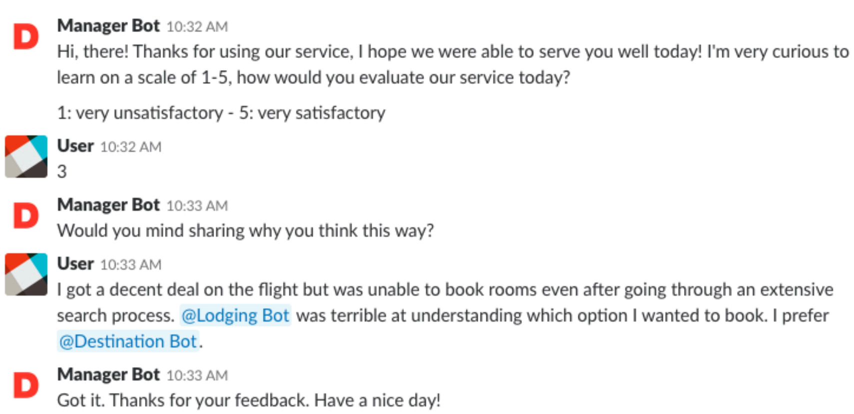


Fig 46. A participant describes the bot-to-service scenario as an extensive process.

Participants tried to communicate without adjusting their language/wording to the bot. While some had success, others realized that it could not understand them, and then they lowered their expectations. Some participants assumed that the bot would not understand them correctly. Some expected to know why the bot did not understand them.

Participants were not entirely confident about a bot that can do all of the multiple tasks in a single experience. A participant said that she tends to expect more from chatbots that they can handle, and she does not also want to "lowball" them as if they are smart. Even the name of the agent, destination bot, created a higher expectation of being able to help with more parts of the travel planning experience for some.

Some participants wished the bot was context-aware. Participants expect the bot to remember their identity, their previous interactions, and their shared history.

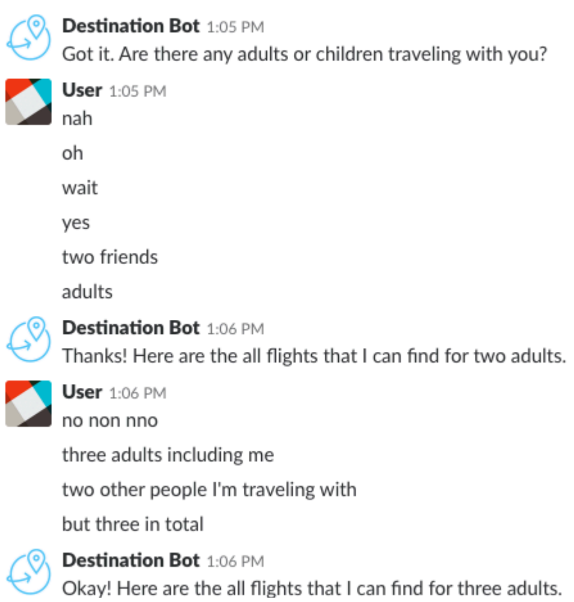


Fig 47. Modifications on conversations were necessary.

Participants wanted more control on bots decision-making and being able to either fact-check or (if necessary) refine its findings. They tried to ask clarification questions. Participants questioned how the bot is making the decision what to show first.

They expected to see proper labeling like “cheapest,” “quickest” options. Many participants questioned if they were getting a good deal from the stranger bot in the multi-bot experience and wanted to trust the agent, but also being able to verify it.

Many participants wished they have been able to change what they wrote to the bot, adjust their preferences later. Some thought that a menu or visual UI elements would make sense in booking stages, which also suggested by the literature.[112]

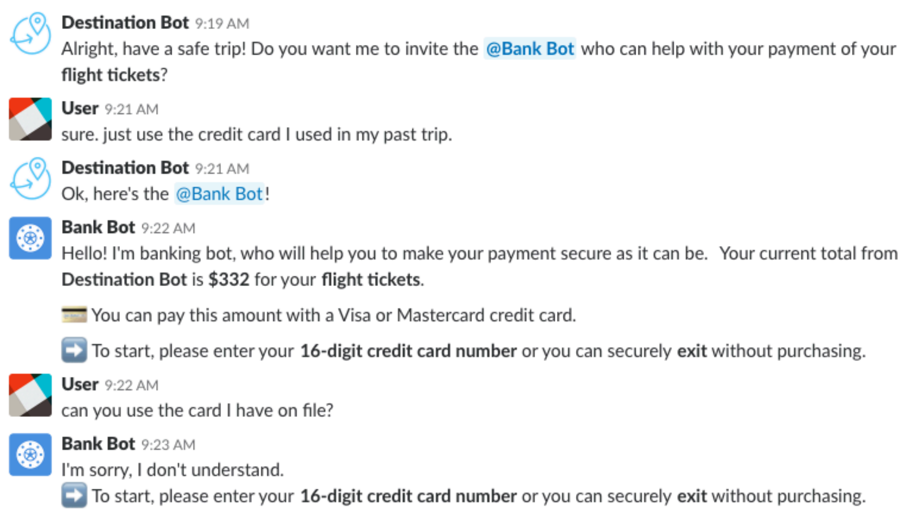


Fig 48. A participant asked agent to use her card on file.

Some participants questioned their trust in the brand of the lodging bot and mentioned they would prefer a branded travel agency to book their lodging. Many participants did not notice bot negotiation; they perceived bots as different websites/brands. Some participants stated that they would trust a banking bot that has a brand image that they know. (i.e., Paypal bot)

[112] Google. “Is Conversation the Right Fit? - Conversation Design Process.” *Google Conversation Design Guideline*, 2018. https:/designguidelines.withgoogle.com/conversation/conversation-design-process/is-conversation-the-right-fit.html.

Timeliness

Participants expected faster response times from all of the bots. Some participants mentioned when bot's response was slower, they would think twice about their interactions with the bot. Similarly, a participant tried to message bots individually through the prior experience with private messaging contacts on Slack when the bot's response had a lag.

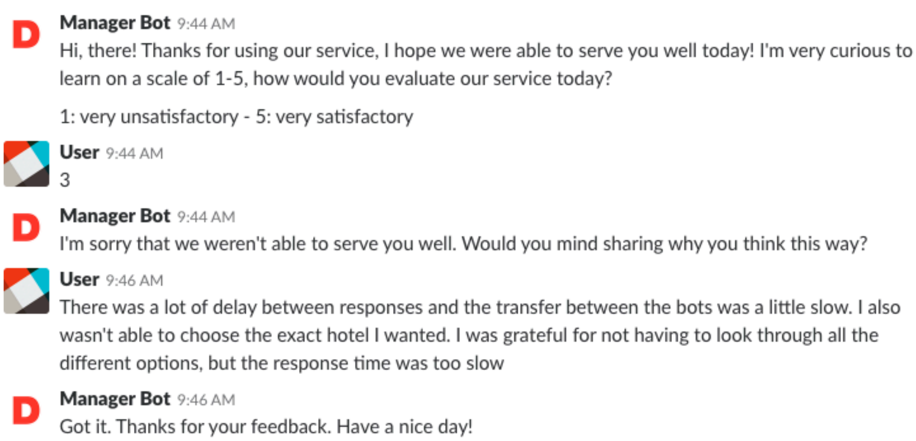


Fig 49. A participant gave explicit feedback about the response rate of the system.

Participants expected to see the status of processing. When they saw a visual cue/indicator, they expressed they are more willing to wait. Seeing a visual cue also helped them to perceive the interaction happening faster. Participants expected bot either reply close to instant or not instant at all to favor multi-tasking. A participant thought that being able to send messages like "this will take awhile, hold on" makes the bot more human and helped to trust it more.

Participants did not expect handed off to another conversational agent. Participants did not expect to interact with multiple bots for the tasks that they did. They assumed that all functions could be handled with a single bot more seamlessly especially if they experienced similar experiences via one channel such as a travel agency website. Some participants found multiple bot approach sophisticated, seamful and nevertheless satisfied with specific bot separations such as banking bot. If the voice and the visual interface is the same, a participant argued that multiple bot system is an illusion of a single bot that changes hats every time.

Participants expected to explain themselves from the start in every handoff because the participants had negative experiences in previous human agent handoffs.

Participants did not expect to approve each handoff while some appreciated how each bot introduces themselves at the beginning of the conversation, which made conversations longer. As participants handed off numerous times, some of them declined to talk with the manager bot because they feel overwhelmed with hand-offs. They also find handoffs slower. Some participants were willing to wait for more if they were interacting with a single bot than multiple bots.

A participant stated that she would not question everything while using such a system if she already trusted the bot. She did not want to see its inner workings such as talking with multiple bots. She did not anticipate to get specific feedback on failed bot-to-bot communication.

A participant interpreted failed communication in the negotiation scenario as if that the stranger bot has "left its shift." He argued that bots are not humans and they do not leave their shifts, and he expected them to be available all the time.

Security, Privacy, Risk

Some participants expect that bots will inevitably make mistakes and therefore they do not trust them with their valuable assets such as money.

Participants were not comfortable giving out their credit card information. They questioned how secure writing sensitive information without visual masks such as asterisks. Some expected to have an external interface (or more private) to input their credit card information (i.e., web-view with SSL).

Specific visual elements increased credibility. Participants find visual parts of the conversation more credible such as search results and emojis.

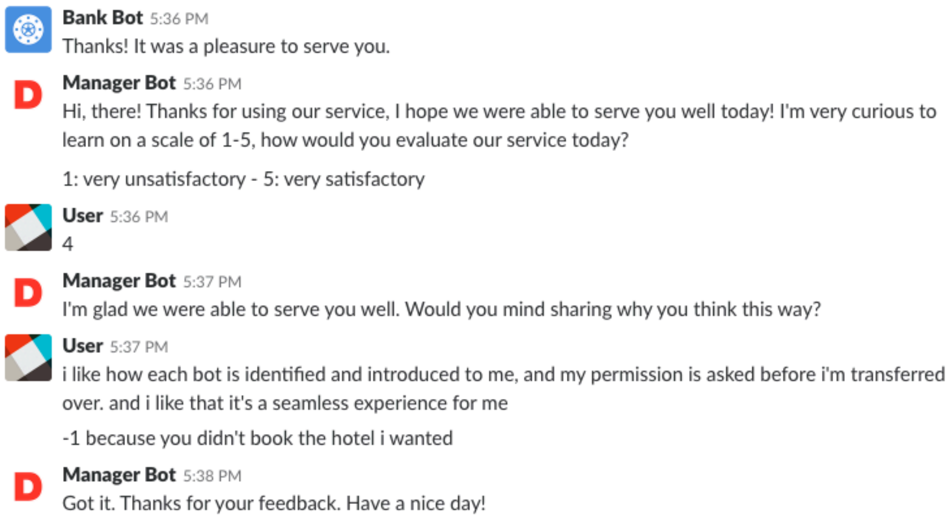


Fig 50. A participant states that he was not expecting to be asked for his consent before sharing his data.

Some participants did not expect a stranger bot to ask for consent to get their data from a trusted bot and found this interaction thoughtful. They assumed that since the interface is the same bots can access their data. Some participants thought it was unnecessary. Some pointed out the discrepancy between how the lodging bot asked for permission and how the bank bot did not ask permission for accessing their data from other bots. Some participants were not sure if other bots were aware of the context of the conversation that they did with other bots.

Generative & Evaluative Research Findings Summary

Generative and evaluative research on travel booking experience showed that participants might not desire handoffs between multiple conversational agents for two main reasons. Hand-offs evoked participants' past negative experiences with human customer care agents. They also made experience feel slower than interacting with a single agent. As agents become seams in an experience, participants did not expect to interact with multiple agents if they were able to accomplish the task with a single actor in the past. In other words, participants did not expect to interact with separate agents when they reserve a hotel, a flight ticket, and then pay it. While agent-collaboration still holds the potential for creating value for participants to accomplish complex tasks, these scenarios should not surface the hand-offs between referrals.

By putting participants deliberately in an unexpected negative situation with multiple agents, this research also shed light on how participants assign responsibility to an agent in a collaboration. In the multiple-agent referral scenario, participants did not blame the trusted agent when the stranger agent made an error, but in the meta-chatbot scenario, participants blame the meta-agent when there was a breakdown, caused by a third party agent.

Discussion

Conversational Trust Design Checklist: Process

Based on my findings on my final design experiment I created a conversational trust checklist for interaction designers in five main categories as a final deliverable. I reviewed existing design checklists to understand 'the best practices for designing best practices.' I, once again, realized how giving an example for each implication helps to memorize and understand its significance.

After I finalized subtitles and descriptions of the implications, I did an online card sorting activity with ten designers, which I asked them to sort all implications free-form, without giving them any categories to put. The card sorting activity helped me how designers understand implications and related them between themselves. Informed by the results of the card sorting and based on my background research findings, I finalized with 14 implications in five categories. I decided to create wireframes to visualize the design suggestions to make them easier to understand. While designing wireframes, I followed Apple Human Interface Guidelines and used my notes about my earlier review on travel chatbots. I aimed to overlay design suggestions on interfaces that designers are already familiar rather than creating a novel interface. I tested the wireframes with five users and iterated on them. Finally, I created short micro interaction videos out of my high fidelity wireframes to leverage the motion to explain specific concepts better such as responsiveness. The high fidelity screen mock-ups is included in the first chapter of this document.

Be Transparent

- ☐ **Share What Agents (Need to) Know About User:** While some expect that every party in a multi-party conversation can access their data, users should get information about their data usage and what all parties know about them.
- ☐ **Refer Others Cautiously, Visualize Confidence Level:** A conversational agent should not refer others (agents or websites) if it is not confident that they can handle the task. Communicate uncertainty with an indicator.
- ☐ **Give Specific Feedback to Clarify**: When there may be a risk for the user such as confirming before a payment, provide detailed and specific feedback to be transparent.

Give Control to the User

- ☐ **Enable Users to Review Bot's Decision-Making:** Communicate the reasoning behind agents' actions and recommendations. Provide a way for users where they can fact-check bot's suggestions and decision-making.
- ☐ **Provide a Room for Revisions:** Users may want to change or update information that they provide to the agent, enable them to do it efficiently.
- ☐ **Fail Gracefully, Offer Auto-Recovery:** In case of failure, provide a reason and a safe exit after two times not to lose the user. Try to do the failed task later, automatically.
- ☐ **Provide Alternatives for Agents:** Some users will not be comfortable with chatting a bot for their high stake transactions yet. Don't be prescriptive, provide alternatives.

Be Relevant

- ☐ **Set the Expectations:** Clearly state what can a bot do or not, how well can it understand the user to eliminate communication breakdowns. The name of the agent can also affect people's expectations.
- ☐ **Remember the Context and Forget it When Asked:** Build upon the previous bits of the conversation. Provide users a sense of memory and a way to forget if needed.

Be Responsive

- ☐ **Indicate the Writing and Processing Visually:** Users expect to see a status of what the bot is doing. They expect to get an answer from a virtual agent quicker than a human. Late responses raise questions about its reliability. A visual indicator that shows whether the bot is writing or processing makes users to perceive bot to be more human and the interaction faster, even it is longer.
- ☐ **Don't indicate Hand-offs:** Don't make the user feel any interruptions and try not to surface the seams in the conversation. Don't emphasize or humanize the hand-offs. Be concise about the first introduction in a hand-off and connect it back to the conversation.

Be Visual

- ☐ **Use visual elements to increase the credibility:** Relevant visual elements tend to increase the trustworthiness of a text-based interface.
- ☐ **Include Branding Where Possible:** To form credibility and show competence, include visual brand symbols such as logos if possible.
- ☐ **Provide Secure Gateways:** Users expect to put their payment information in secure and encrypted forms on a cognitively higher level than the conversation. Leverage solutions that can show the security level of the transaction such as a webview with https:/ page.

Limitations

This project included design experiments and qualitative analysis that involves subjective views and anecdotal findings; my judgments may have introduced a bias in findings and implications.

Limitations due to the external sources most significantly relatable to the representation of the participants that I included in this study. My primary user group for my design experiments was university students around the same age. Although their schools had an international representation, the majority of the participants had a background in design. While all studies had a balanced number of male and females, the origin of the participants was unbalanced. For these reasons, my implications and results from this study cannot be directly generalized, and more research/testing should be performed. Also, the small number of participants prevent the quantitative analysis of research from providing conclusive results. Therefore, implications and findings from this project are suggestions for designers, rather than definite facts.

This thesis also included Wizard of Oz experiment to test my assumptions faster and eliminate the limitations of natural language processing. Nevertheless, it was not possible to simulate the processing speed of a computer, which decreased the response time of chatbots in the experiment. As the responsiveness of the system associated with its reliability, it affected how trustworthy were the bots perceived. Therefore I included "be responsive" as one of my design implications for more trustworthy conversational agents. Since I asked participants to role-play in this experiment, their actions may have also influenced by several experimental effects such as participant reactivity and novelty effect.

Future Work

Doing More User-Research for Evaluation

To tackle some threats discussed "issues around validity" section, I plan more user tests for testing my findings and more subject-matter expert interviews to get their opinion and critique on the relevancy of the implications. On the other hand, creating an iteration of my Wizard of Oz research with real chatbots would help significantly to locate additional limitations of Wizard of Oz approach and how it influenced my findings. Also, while I was able to do a pilot usability test of the wireframes of my implication examples with five users, I should do more testing to fine tune my design proposals.

Making Implications More Useful, Approachable and Actionable

Due to the timespan of the project, I was only able to build and test, yet not entirely thought about how to present the value of the conversational trust design checklist in the most useful and actionable way for designers and others who are interested. As mentioned in the final design experiment reflection, I believe presenting this work in the most inspiring way would make those who are interested in to think about the trust and conversational agents more concretely and actionable. While I worked on creating tangible examples for each of my implications, I believe I can also leverage other media such as an article on Medium.com or an explainer-video to make things more glanceable and manageable by designers with busy minds and less time.

Implications for Different Modalities

While my implications aimed for text chatbots, I believe they also apply in different interaction modalities such as voice-based agents or embodied agents. The guidelines can be used as a starting point for designers in these areas to explore trust and collaboration in conversational agents. Additional research would be required to test which implications would need adaptation to the modality of the conversation.

Blame, Responsibility Research in Conversational Agents

Findings around the relationship between blame and responsibility unveiled a link to trust repair with conversational agents. While in this study conversational agents did not try to repair trust and blame the user as suggested in the literature[113], agents blamed a third party for preventing loss of trust in a case of an error.

Differences between Algorithmic Conversational Experiences

If we think conversational agents as black-box computer programs, I believe this project would be also helpful for someone designing not only a chatbot, but also algorithms. As trusting an agent may mean trusting an algorithm, I believe there is a significant overlap between two concepts. Further research would be needed to explore the nuances of the 'algorithmic experience'.

Agent Collaboration in Different Contexts

In this thesis, I was able to position trust and collaboration in an e-commerce setting, more specifically on a travel booking scenario. I am also curious to see how would be the implications defined for different scenarios in e-commerce and more importantly the use of agent-collaboration in different problem spaces such as healthcare, mission-critical services, and finance.

[113] Google. "Apologies - Conversational Components." *Google Conversation Design Guideline*, 2018. https:/designguidelines.withgoogle.com/conversation/conversational-components/apologies.html.

Reflection

Trust is complex.

Working between high-level design strategies and architectures to granular visual and conversation design decisions made me realize how vital, yet complex trust is for establishing and maintaining the relationship between humans and technology artifacts.

Conversations are for building trust.

Combining trust and conversation into a single model taught me how ‘building’ trust is parallel with conversing. In other words, I learned how trust becomes the outcome of a conversation.

Trust is contextual.

Seeing many trust definitions in the literature pushed me to scope my research better. While I trusted the nature of the process that took me where I finalized, I learned that I had to make decisions my research directions earlier for similar projects in future.

Just enough research is what is necessary.

The short timespan of thesis taught me how much primary and secondary research would be necessary to advance and also to research using my skills in design and making.

Trust is going to be more important in future and I am just starting...

Having this thesis project as a foundation to my future work in trust, I believe I still, have a lot to learn, test, and verify as a researcher and designer to design for int eractions that will users trust and start a conversation.

Conclusion

In this thesis project, I explored how designers can design for trust and collaboration in conversational agents, taking the e-commerce domain as an example. Throughout the background research, user research, and design experiments, I learned about the nuances of trust and conversation design. Scoping down to a fragmented experience, a travel booking journey, I tested a Wizard of Oz prototype for two collaboration scenarios in order to compare how trust dynamics changes, accordingly. Using my findings, I designed a conversational trust checklist to illustrate the design implications.

My findings from Wizard of Oz prototype suggested that users do not expect to interact with different agents. They do not expect to be handed-off between agents to do separate tasks if they can do the same tasks through a single agent. As users experienced a staged breakdown in Wizard of Oz prototype, I found out that in case of a breakdown with a stranger agent, users do not lose their trust to the advisor (trusted) agent. Also, I found out that users blame the stranger agent in a multi-agent scenario if the stranger agent let them down, rather than the advisor agent. In contrast, in a negotiation scenario, if a trusted agent has an issue with a stranger agent and blames stranger agent for the error, users blamed trusted agent, not the stranger agent.

To discuss my findings and create suggestions for interaction designers, who are interested in designing for trust, I finalized my thesis with 14 implications in five categories: be transparent, give control to the user, be relevant, be responsive, and be visual.

Legal

This thesis does not seek in any way to profit commercially. This thesis is intended as an example of the owner's research and scholarship. This thesis is not sponsored or approved by any third-party business or company.

All third-party content that have been used in this thesis are transformative works used in good faith, in keeping with Section 107 of U.S. copyright law, and as such constitute fair use of copyrighted material. All rights of these content belong to their legal owners.

According to Section 107 of U.S. copyright law, "the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright." The copyrighted materials included on this thesis are used for all of these purposes, with the exception of teaching and news reporting.

This thesis uses copyrighted material for illustration purposes. Using copyrighted examples for Illustration purposes is necessary when commenting on and critiquing the existing literature related to trust, conversational agents, and collaboration. To represent real-life data

To discuss fair use with the owner of this thesis, please email mericda@gmail.com.

Bibliography

Gaunt, Kevin. "Living with Bots - Battling Boredom in Smart Homes," 2015. https://www.kevingaunt.com/.

Juez, Laura Alba. "Verbal Irony and the Maxims of Grice's Cooperative Principle." *Revista Alicantina de Estudios Ingleses* 8, no. 8 (1995): 25–30.

Luger, Ewa, and Abigail Sellen. "'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents." *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*, 2016, 5286–97.

Krishna, Golden. *The Best Interface Is No Interface : The Simple Path to Brilliant Technology*, n.d.

Nass, Clifford, Jonathan Steuer, and Ellen R. Tauber. "Computers Are Social Actors." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence - CHI '94*, 72–78. New York, New York, USA: ACM Press, 1994.

Coleman, James S. *Foundations of Social Theory*. Belknap Press of Harvard University Press, 1990.

Lewis, Mike, Denis Yarats, Yann Dauphin N., Devi Parikh, and Dhruv Batra. "Deal or No Deal? Training AI Bots to Negotiate." *Engineering Blog, Facebook Code*, 2017. https://code.facebook.com/posts/1686672014972296/deal-or-no-deal-training-ai-bots-to-negotiate/.

Chalmers, Matthew, and Ian Maccoll. "Seamful and Seamless Design in Ubiquitous Computing." Accessed May 12, 2018. http://www.techkwondo.com/external/pdf/reports/2003-chalmers.pdf.

Turing, Alan M. "Computing Machinery and Intelligence." In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, 49:23–65, 2009.

Knight, F. H. "Risk, Uncertainty and Profit, 1921." *Boston and New York* 23, no. 3 (1921): 135–57.

Google. "Conversation Design." *Google Design Guidelines*, 2018. https://designguidelines.withgoogle.com/conversation/#.

Chaves, Ana Paula, and Marco Aurelio Gerosa. "Single or Multiple Conversational Agents? An Interactional Coherence Comparison." *Chi 2018*, no. February (2018).

Bhattacharya, Rajeev, Timothy M. Devinney, and Madan M. Pillutla. "A Formal Model of Trust Based on

Outcomes." *The Academy of Management Review* 23, no. 3 (July 1998): 459.

Eltes, Jonas. "Lost in Computation," 2017. https://vimeo.com/204531768.

O'Hara, Kieron, Harith Alani, Yannis Kalfoglou, and Nigel Shadbolt. "Trust Strategies for the Semantic Web." *Proceedings of the 2004 International Conference on Trust, Security, and Reputation on the Semantic Web - Volume 127*. CEUR-WS.org, 2004.

Interbot. "Bot to Bot Communication Platform." *Interbot*, 2016. https://www.interbot.cc/.

Google. "Is Conversation the Right Fit? - Conversation Design Process." *Google Conversation Design Guideline*, 2018. https://designguidelines.withgoogle.com/conversation/conversation-design-process/is-conversation-the-right-fit.html.

Cleverbot. "Clever Them! Cleverbot AI Avatars Talk to Each Other about Life, the Universe and Everything." Accessed May 12, 2018. http://www.cleverbot.com/cleverthem.

"Conversational Design Guidelines - Google Search," 2018. https://www.google.com/search?q=conversational+design+guidelines.

Greater than Experience Design Consultancy. *Designing for Trust: The Data Transparency Playbook*. 2nd ed., 2018. https://www.greaterthanexperience.design/getplaybook/designing-for-trust-the-playbook.

Millard, Miss. "Grice's Maxims in 'The Big Bang Theory.'" *Youtube*, 2014. https://www.youtube.com/watch?v=vEM8gZCWQ2w.

Candello, Heloisa, and Bruna Daniele Andrade. "Evaluating Multi-Agent Conversational Interfaces in the Early Stages of the Design Process" 1 (2016): 1–15.

Krause, Elliot. "This Robot Will Handle Your Divorce Free of Charge." *The Wall Street Journal*, 2017. https://www.wsj.com/articles/this-robot-will-handle-your-divorce-free-of-charge-1522341583.

Gibbs, Samuel. "Elon Musk: Artificial Intelligence Is Our Biggest Existential Threat." *The Guardian*, 2014. https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat.

Weizenbaum, Joseph. "ELIZA — A Computer Program For the Study of Natural Language Communication Between Man And Machine." *Communications of the ACM* 9, no. 1 (1966): 36–45.

Farokhmanesh, Megan. "Deepfakes Are Disappearing from Parts of the Web, but They're Not Going Away." *The Verge*, 2018. https://www.theverge.com/2018/2/9/16986602/

de Visser, Ewart J., Richard Pak, and Tyler H. Shaw. "From 'Automation' to 'Autonomy': The Importance of Trust Repair in Human–Machine Interaction." *Ergonomics* 0139 (2018): 1–19.

Majd, Elham, and Vimala Balakrishnan. "A Trust Model for Recommender Agent Systems." *Soft Computing* 21, no. 2 (January 25, 2017): 417–33.

Marge, Matthew, João Miranda, Alan W Black, and Alexander I Rudnicky. "Towards Improving the Naturalness of Social Conversations with Dialogue Systems." In *Proceedings of SIGDIAL 2010: The 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 91–94, 2010.

Bloomberg, Jason. "Should AI Fool You? Think Again." *Intellyx*, 2017. https:/intellyx.com/2017/10/23/should-ai-fool-you-think-again/.

"Conversation Design Quick Reference Write Sample Dialogs," 2017, 2017.

Sauer, Ben. "Voice Guidelines." *Clearleft*, 2017. https:/voiceguidelines.clearleft.com/.

Wolffe, Richard. "The Evil Genius of Cambridge Analytica Was to Exploit Those We Trust Most." *The Guardian*, 2018. https:/www.theguardian.com/uk-news/2018/mar/21/cambridge-analytica-facebook-exploited-trust.

White, Mathew P. "The Trust Paradox: The Role of Context Effects in Stated Trust Judgements." *Journal of Risk Research* 10, no. 7 (October 2007): 977–88.

Lewis, Mike, Denis Yarats, Yann N. Dauphin, Devi Parikh, and Dhruv Batra. "Deal or No Deal? End-to-End Learning for Negotiation Dialogues," 2017.

Maass, Wolfgang, and Upkar Varshney. "Preface to the Focus Theme Section: 'Smart Products.'" *Electronic Markets* 18, no. 3 (August 2008): 211–15.

Leviathan, Yaniv, and Matias Yossi. "Google AI Blog: Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone." *Google AI Blog*, 2018. https:/ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

Pangaro, Paul. "Conversation Is More than Interface," no. February (2017).

Warren, Tom. "Microsoft Demonstrates Alexa and Cortana Integration." *The Verge*, 2018. https:/www.theverge.com/2018/5/7/17326808/microsoft-cortana-update-alexa-integration-build-2018.

Botsman, Rachel. *Who Can You Trust? : How Technology Brought Us Together : And Why It Could Drive Us Apart*, 2017.

Cohen, Michael H., James Giangola, and Jennifer Balogh. *Voice User Interface Design*, 2004.

Google. “Implicit Invocation.” *Actions on Google*, 2018. https://developers.google.com/actions/discovery/implicit.

Beer, Austin. “Let’s Chat!” Hyper Island, 2016.

Superflux. “Our Friends Electric,” 2017. http://superflux.in/index.php/work/friends-electric/#.

Medhi Thies, Indrani, Nandita Menon, Sneha Magapu, Manisha Subramony, and Jacki O’Neill. “How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians.” In *Lecture Notes in Computer Science*, 10513 LNCS:441–59, 2017..

Riegelsberger, Jens, M. Angela Sasse, and John D. McCarthy. “The Mechanics of Trust: A Framework for Research and Design.” *International Journal of Human-Computer Studies* 62, no. 3 (March 1, 2005): 381–422.

Google. “Apologies - Conversational Components.” *Google Conversation Design Guideline*, 2018. https://designguidelines.withgoogle.com/conversation/conversational-components/apologies.html.

Byrne, Daragh. “In-Class Exercise II: Personalities.” *Programming for Online Prototypes*, 2016. https://daraghbyrne.github.io/onlineprototypes2016/in-class/in-class-2/.

Botchain. “Can You Trust AI?,” 2018. https://botchain.talla.com/.

Giangola, James. *Applying Built-in Hacks of Conversation to Your Voice UI (Google I/O ’17)*. Youtube, 2017. https://www.youtube.com/watch?v=wuDP_eygsvs.

McKnight, D. Harrison, Nancy K. Lankton, Andreas Nicolaou, and Jean Price. “Distinguishing the Effects of B2B Information Quality, System Quality, and Service Outcome Quality on Trust and Distrust.” *Journal of Strategic Information Systems* 26, no. 2 (2017): 118–41.

Fakourfar, Omid, and Benny Lin. “The Future of Mobile Travel Booking: Travel Chatbots: Are We There Yet ?,” 2017.

Google. “In Conversation, There Are No Errors.” Youtube, 2017. https://www.youtube.com/watch?v=oOLo071Pj1U.

Carbon Healthcare. “A Fresh Take on Health Care,” 2017. https://carbonhealth.com/ehr.

Amazon. “About Timers & Alarms.” *Amazon.Com Help*. Accessed May 9, 2018. https://www.amazon.com/gp/help/customer/display.html?nodeId=201549770.

Geniola, Alberto, Markku Antikainen, and Tuomas Aura. “Automated Analysis of Freeware Installers Promoted by Download Portals.” *Computers & Security* 77 (August 1, 2018): 209–25.

Sure. “Chatbot for Restaurant Recommendations on Messenger.” Accessed May 13, 2018. http://surebot.io/.

Yuksel, Beste F, Penny Collisson, and Mary Czerwinski. “Brains or Beauty: How to Engender Trust in User-Agent Interactions.” *ACM Transactions on Internet Technology (TOIT)* 17, no. 1 (2017): 1–20.

Google. “Create a Persona - Conversation Design Process.” *Google Conversation Design Guideline*, 2018. https://designguidelines.withgoogle.com/conversation/conversation-design-process/create-a-persona.html.

Pearl, Cathy. *Designing Voice User Interfaces Principles of Conversational Experiences*, 2016. https://books.google.cz/books?id=WGnEDQAAQBAJ.

Google. “Explicit Invocation.” *Actions on Google*, 2017. https://developers.google.com/actions/discovery/explicit.

Recouvreux, Thomas, and Jo Colina. “IDK Protocol - The “I Don’t Know Protocol” for Chatbots.” Accessed May 12, 2018. http://idkproto.com/.

Shevat, Amir. “5 Scenarios for How Humans and Bots Work Together.” *Slack Platform Blog*, 2016. https://medium.com/slack-developer-blog/5-scenarios-for-how-humans-and-bots-work-together-a0e636053ef4.

Google. “Errors - Conversational Components.” *Google Conversation Design Guideline*, 2018. https://designguidelines.withgoogle.com/conversation/conversational-components/errors.html.

Dubberly, Hugh, and P Pangaro. “What Is Conversation? How Can We Design for Effective Conversation?” *Interactions Magazine* XVI, no. August (2009): 1–9.

Data, The, and Transparency Playbook. “Designing for Trust The Data Transparency Playbook,” 2017.

Grice, Paul. “Logic and Conversation.” *Syntax and Semantics*, 1975.

Candello, Heloisa, and Bruna Daniele Andrade. “Evaluating Multi-Agent Conversational Interfaces in the Early Stages of the Design Process” 1 (2016): 1–15.

Nass, Clifford, and Scott Brave. *Wired for Speech: How Voice Activates and Advances the Human Computer Relationship*. *Computational Linguistics*. Vol. 32. MIT, 2005.

Pangaro, Paul. “Conversation Is More than Interface.” *Pangaro.Com*, 2017. http://pangaro.com/ixda2017/index.html.

Mayer, Roger C, James H Davis, and F David Schoorman. “An Integrative Model of Organizational

Trust.” *Source: The Academy of Management Review* 20, no. 3 (1995): 709–34.

Datta, Saurabh. “BOT Panel Evolution.” *Saurabh Datta*, 2017. http://www.dattasaurabh.com/BOT-Panel-evolution.

TED. “Joe Gebbia: How Airbnb Designs for Trust | TED Talk.” *TED*, 2016. https://www.ted.com/talks/joe_gebbia_how_airbnb_designs_for_trust.

Google. “Test and Iterate - Conversation Design Process.” *Google Conversation Design Guideline*, 2018. https://designguidelines.withgoogle.com/conversation/conversation-design-process/test-iterate.html.

Mcknight, Dh, M Carter, Jb Thatcher, and Pf Clay. “Trust in a Specific Technology: An Investigation in Its Components and Measures.” *ACM Transactions on Management Information Systems* 2, no. 2 (2011): 1–12.

Knight, Will. “Can This Man Make AI More Human? - MIT Technology Review.” *MIT Technology Review*, 2015. https://www.technologyreview.com/s/544606/can-this-man-make-ai-more-human/.

if design consultancy. “Trust & Design Meetup.” Accessed May 12, 2018. https://trustanddesign.projectsbyif.com/.

Aboluwarin, Pelumi. “Chatbots — Igniting Division of Labour in AI.” *Medium*, 2016. https://medium.com/@pelumi/chatbots-igniting-division-of-labour-in-ai-1430fcc85c8d.

Lankton, Nancy, D. Harrison McKnight, and John Tripp. “Technology, Humanness, and Trust: Rethinking Trust in Technology.” *Journal of the Association for Information Systems* 16, no. 10 (2015): 880–918.

Glaser, Judith E. *Conversational Intelligence : How Great Leaders Build Trust and Get Extraordinary Results*, n.d.

Pollina, Dean A., and Allison Barretta. “The Effectiveness of a National Security Screening Interview Conducted by a Computer-Generated Agent.” *Computers in Human Behavior* 39 (October 1, 2014): 39–50.

Cherian, Elizabeth, and Jeremy Pounder. “Speak Easy Global: The Futures Answers to You,” 2017.

Jøsang, Audun, and Simon Pope. “Semantic Constraints for Trust Transitivity.” In *Conferences in Research and Practice in Information Technology Series*, 43:59–68, 2005.

Endnotes

[A] Johnson, Khari. "Digit Bot Has Saved People $250 Million and Is Now Available on Facebook Messenger." VentureBeat, 2017. https:/venturebeat.com/2017/01/10/digit-bot-has-saved-people-250-million-and-is-now-available-on-facebook-messenger/.

[B] Woebot. "Woebot - Your Charming Robot Friend Who Is Ready to Listen, 24/7." Accessed May 15, 2018. https:/woebot.io/.

[C] "DoNotPay." Accessed May 15, 2018. https:/www.donotpay.com/.

[D] "DoNotPay." Accessed May 15, 2018. https:/www.donotpay.com/.

[E] Icelandair. "Book Your Flight with Icelandair's New Messenger Bot." Icelandair, 2016. https:/www.icelandair.com/blog/messenger-bot/.

[F] Johnson, Khari. "Digit Bot Has Saved People $250 Million and Is Now Available on Facebook Messenger." VentureBeat, 2017. https:/venturebeat.com/2017/01/10/digit-bot-has-saved-people-250-million-and-is-now-available-on-facebook-messenger/.

[G] Vincent James. "AI That Detects Cardiac Arrests during Emergency Calls Will Be Tested across Europe This Summer." The Verge, 2018. https:/www.theverge.com/2018/4/25/17278994/ai-cardiac-arrest-corti-emergency-call-response.

[H] Google. "Actions on Google." Google Developers. Accessed May 12, 2018. https:/developers.google.com/actions/.

[I] Amazon. "Design Checklist." Amazon Alexa Voice Design Guide. Accessed May 11, 2018. https:/developer.amazon.com/designing-for-voice/design-checklist/.

[J] Microsoft. "Cortana Skills." Microsoft Cortana. Accessed May 12, 2018. https:/www.microsoft.com/en-us/cortana/skills.

[K] Apple. "SiriKit." Apple Developer Documentation. Accessed May 12, 2018. https:/developer.apple.com/documentation/sirikit.

L Facebook. "Messenger Bots," Accessed May 15, 2018. https:/messenger.fb.com/.

[M] Google. "Conversation Design." Google Design Guidelines, 2018. https:/designguidelines.withgoogle.com/conversation/#.

[N] Warren, Tom. "Microsoft Demonstrates Alexa and Cortana Integration." The Verge, 2018. https:/

www.theverge.com/2018/5/7/17326808/microsoft-cortana-update-alexa-integration-build-2018.

O Leviathan, Yaniv, and Matias Yossi. “Google AI Blog: Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone.” Google AI Blog, 2018. https:/ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html.

P Amazon. “Design Checklist.” Amazon Alexa Voice Design Guide. Accessed May 11, 2018. https:/developer.amazon.com/designing-for-voice/design-checklist/.

Q Apple. “Siri - System Capabilities - IOS Human Interface Guidelines.” Accessed May 15, 2018. https:/developer.apple.com/ios/human-interface-guidelines/system-capabilities/siri/.

R Moore, Bob, and Raphael Arar. “Introduction.” IBM Conversational UX. Accessed May 11, 2018. https:/conversational-ux.mybluemix.net/design/conversational-ux/.

S Microsoft. “Cortana Skills.” Microsoft Cortana. Accessed May 12, 2018. https:/www.microsoft.com/en-us/cortana/skills.

T Beer, Austin. “How to Design Intelligence: 3 Ways to Make Human-Centered Bots.” The Charming Device, 2016. https:/medium.com/the-charming-device/how-to-design-intelligence-3-ways-to-make-human-centered-bots-76c5ff7524df.

U “Walkie - Bots Gone Wild.” Accessed May 15, 2018. https:/walkiebot.co/.

V “Dialogflow.” Accessed May 15, 2018. https:/dialogflow.com/.

W “Design, Preview and Prototype Your next Chatbot or Voice Assistant.” Accessed May 15, 2018. https:/botsociety.io/.

X “Walkie - Bots Gone Wild.” Accessed May 15, 2018. https:/walkiebot.co/.

Y “Create a Facebook AI Chatbot Without Coding.” Accessed May 15, 2018. https:/chatfuel.com/.

Z Google. “Conversation Design.” Google Design Guidelines, 2018. https:/designguidelines.withgoogle.com/conversation/#