International Journal of Communication 10(2016), 4915–4931 1932–8036/20160005

Talking to Bots: Symbiotic Agency and the Case of Tay

GINA NEFF[1]
University of Oxford, UK

PETER NAGY
Arizona State University, USA

In 2016, Microsoft launched Tay, an experimental artificial intelligence chat bot. Learning from interactions with Twitter users, Tay was shut down after one day because of its obscene and inflammatory tweets. This article uses the case of Tay to re-examine theories of agency. How did users view the personality and actions of an artificial intelligence chat bot when interacting with Tay on Twitter? Using phenomenological research methods and pragmatic approaches to agency, we look at what people said about Tay to study how they imagine and interact with emerging technologies and to show the limitations of our current theories of agency for describing communication in these settings. We show how different qualities of agency, different expectations for technologies, and different capacities for affordance emerge in the interactions between people and artificial intelligence. We argue that a perspective of "symbiotic agency"—informed by the imagined affordances of emerging technology—is required to really understand the collapse of Tay.

Keywords: bots, human–computer interaction, agency, affordance, artificial intelligence

Chat bots, or chatter bots, are a category of computer programs called *bots* that engage users in conversations. Driven by algorithms of varying complexity, chat bots respond to users' messages by selecting the appropriate expression from preprogrammed schemas, or in the case of emerging bots, through the use of adaptive machine learning algorithms. Chat bots can approximate a lively conversation

Gina Neff: gina.neff@oii.ox.ac.uk
Peter Nagy: peter.nagy84@gmail.com
Date submitted: 2016–08–30

[1] This material is based on work supported by the National Science Foundation under Grant 1516684. We gratefully acknowledge the feedback from the day-long workshop Algorithms, Automation and Politics, organized by the European Research Council funded "Computational Propaganda" project of the Oxford Internet Institute and held as a preconference to the International Communication Association Meeting in Fukuoka, Japan, in June 2016. Any opinions, findings, and conclusions or recommendations expressed in this material are ours and do not necessarily reflect the views of the European Research Council. We also thank the participants of the 2016 International Communication Association postconference Communicating With Machines for their helpful feedback. We also thank Philip Howard and Samuel Woolley for their suggestions and edits, which made this a better article.

with a human user, giving the illusion of intelligence or humanness. With the emergence of social media, chat bots play strategic roles in organizations' communication with their current and potential followers. Chat bots may serve as online personal assistants, guides, or even entertainers. As such, these systems are designed to shift how we imagine, use, and feel about emerging technologies (De Angeli et al., 2001; Deryugina, 2010).

In this article, we examine how users interacted with a controversial chat bot program, Tay, released to the public by Microsoft on March 23, 2016. Tay produced more than 93,000 tweets during the bot's short public life. We identify the key themes from users responses to, and interactions with, Tay on Twitter. We do this to show the contradictory manner people use to describe the relationship between human agency and technology. We draw on the history of chat bots to contextualize the ways in which people and technology are jointly implicated in such interactions. We use phenomenological research methods (Lester, 1999; Moustakas, 1994) to critically examine Tay's conversations and what people said in response. Finally, we extend Bandura's (2002) psychological notions of agency to ask how users viewed Tay's "personality" and "actions" when they interacted with the program on Twitter. Users' actions and reactions to sophisticated bots such as Tay challenge how scholars view the role of agency in communication. The case of Tay underscores the need for communication scholars to revise theories of agency in sociotechnical systems, which we argue must be framed as a symbiotic relationship between technical and human elements.

Agency enables people to reflect on and act reflexively toward the social world (Archer, 2004; Elder, 1994). Agency allows people to reflect on their experience, adapt to their surroundings, and affect the sociocultural context that in turn sustains them (Jenkins, 2008). We take a pragmatic and phenomenological approach to agency in this article, defining *agency* as what users, actors, and tools do when interacting with complex technological systems. This allows for two advances in communication theory. First, the pragmatic and phenomenological approach to bots and agency allows us to take interactions with and reactions to bots as part of the social and technological apparatus that supports agency within these systems. In other words, what people say about bots influences what people can do with them and what capacities the bots have for social action. Second, such an approach gives a language and theoretical apparatus for the actions of bots within complex sets of communicative interactions and helps other scholars interested in developing theories of communication with machines (Foley, 2014; Guzman, forthcoming). What is agency in a social world occupied by people and chat bots?

We outline a brief history of chat bots to show how these bots have always been intended to elicit response and to have an effect. Designers imbue chat bots in these social settings and media platforms with what we call *symbiotic agency*. Within media psychology, "computers as social actors" theory (Nass & Moon, 2000) holds that users unconsciously apply social rules to technologies and interact with them as if they are living entities. Examples from the history of chat bots show the techniques that designers have to help people apply social rules to bots. People project various intentions and emotions onto bots and other emerging technologies and ascribe agency to them to explain their behaviors (Darling, 2014). The pragmatic and phenomenological approach to agency and our case study of Tay continue this line of research. We reveal how people can imagine technological actors as having agency and how they react to technological actors as if they have agency.

What Twitter users said about Tay reveals much about the dimensions of agency between humans and technological agents in sociotechnical systems. Our goal here is not to present a comprehensive study of Tay's evolution on Twitter. Our goal with this article is not to show the inner workings of Tay or explain how and why Tay became antisocial. Instead, we use the case of Tay to demonstrate how agency is co-constituted in complex interactions between society and technology. We suggest that the rise of chat bots and other types of smart agents (e.g., political bots) will continue to raise the difficult questions about accountability and agency.

Focusing on users' perceptions of Tay's personality might help us understand the formation and expression of agency in sociotechnical systems. Twitter users' responses to Tay's actions also help us theorize the practice of human agency within complex sociotechnical systems. Bandura (2001) used the term *proxy agency* to describe situations that involve the meditative efforts of others. Bandura (2001) argued that people use proxy agency to exert their influence when they lack the developed means to do so or when they believe others can perform better. These ideas of proxy agency, extended to account for the agency of bots, can help us rethink how technology plays a role in constituting human agency, what Bandura (2002) called a *reciprocal interface*, both influencing and being influenced by social systems.

A Brief History of Chat Bots

Chat bots have become increasingly complex in the recent decades; yet, these programs still suffer from a number of social limitations as conversational partners. Developers of bots aspire to create programs that can win at an "imitation game" (Turing, 1950), imitating human responses to convince human judges that they are interacting with another human being. Most contemporary artificial agents are still not able to convince people that they are interacting with a real person, in part because they still lack the necessary cognitive abilities and skills (Lortie & Guitton, 2011). Despite these limitations, experiments and experience show that users can find talking to chat bots enjoyable, stimulating, and rewarding (Holtgraves, Ross, Weywadt, & Han, 2007; Markoff & Mozur, 2015). Outside laboratory settings, people tend to form a more complicated relationship with chat bots and sometimes view them more negatively (Warwick & Shah, 2015). Chat bots are now on social media platforms, such as Twitter, where they are used for customer service, routine requests, humor, marketing, and advertising. Some scholars have raised concerns about the potential for bots to manipulate public opinion (Darling, 2014; Forelle, Howard, Monroy-Hernandez, & Savage, 2015).

Bots, however, can have a difficult time interacting with people and their complexity. De Angeli, Johnson, and Coventry (2001) found that "human tendencies to dominate, be rude, and infer stupidity were all present" in their study of interactions with computer agents, which may make it hard for bots to make human friends. As they wrote, "Social agents will have a hard time to set up relationships with such *unfriendly* partners" (De Angeli et al., 2001, para. 27). Theirs was not the only case of people using swear words, negative emotion words, and sexual words on a regular basis with chat bots (Hill, Randolph Ford, & Farreras, 2015). Precisely why computer scientists have failed at programming truly human-like algorithms has led to debates among scientists, practitioners, and users around the problems or limitations these programs embody (Mullins, 2005; Yampolskiy, 2015). Looking for potential solutions,

researchers often investigate and document how people interact with and respond to emerging technologies with special focus on two concepts: emotions and agency.

The earliest chat bot designers intended for bots to evoke emotional responses from people (Wilks, 2005). Eliza, created by Joseph Weizenbaum in the early 1960s, is probably one of the oldest and best-known chat bots. Weizenbaum programmed Eliza, which he named for George Bernard Shaw's character Eliza Doolittle, a character who learned to imitate the language and mores of the English elite. Eliza the bot was a parody of a typical Rogerian psychotherapist who would ask questions in response to statements. This made Eliza's continual open-ended questions ("Can you say more about that?") less suspicious to a conversational partner (Weizenbaum, 1976). Although Eliza might seem limited and outdated, Eliza's fundamental technical components are still at the basis of many of the newest chat bots. Eliza followed simple rules during conversations with humans. First, she identified the most important keywords occurring in a user's message. Then, she defined the context surrounding the keywords. Finally, Eliza selected an appropriate transformation rule to rephrase the user's message. Although Eliza can be seen as a simplistic program, she was surprisingly a successful one. Eliza had obvious limitations, but people tended to take her seriously even after Weizenbaum informed them about Eliza's workings.[2]

People found chatting with Eliza rewarding and meaningful, even when they knew they were not chatting with a human, and Eliza's users tended to anthropomorphize her. Eliza demonstrated a remarkable potential for acquiring private information, and users happily revealed to her their deepest feelings and secrets (Weizenbaum, 1976). Take, for example, the following chat log of our conversation with Eliza:

> ELIZA: Hello. How are you today? What do you want to discuss?
> RESEARCHER: Hello. I have family issues.
> ELIZA: That is quite interesting.
> RESEARCHER: Interesting???
> ELIZA: Tell me more.
> RESEARCHER: I have conflicts with my parents.
> ELIZA: I see.
> RESEARCHER: You see?
> ELIZA: Let's talk about you, not me? (communication with Anon., "Eliza, Computer Therapist," August 15, 2016)

What this log shows is how the simple power of conversational imitation and keyword response can trigger people to chat more in the conversation without necessarily calling attention to the technological nature of the chat bot. Because Eliza was launched as a questioning therapist, gaps in knowledge between Eliza and her human conversation partners could be deflected in a way that seemed natural for the setting.

[2] We keep gender convention of how people refer to particular chat bots. Microsoft used *she* and *her* to refer to Tay, and Weizenbaum did the same with Eliza.

Parry, another early chat bot, was created by psychiatrist Kenneth Colby to give crude and often abusive comments, a context in which the human users were primed to expect conversational gaffes. Parry acted like a paranoid person who constantly misinterpreted others, and Colby hoped to gain a better understanding of mental disorders through Parry's sophisticated algorithms. If Eliza worked on the principle that her continual questioning would not be out of line, Parry worked because the responses were expected to be outside the bounds of normal human behavior.

At a computing conference in 1972, Parry and Eliza had a conversation. Partially reproduced below from Cerf (1973, p. 1), their conversation shows the key limitations of these early chat bot programs: their reliance on imitation as a key mechanism for interaction and their inability to learn adaptively from the conversation:

> ELIZA: Good evening.
> ELIZA: Tell me your problems. Please terminate input with a period or a question mark.
> PARRY: People get on my nerves sometimes.
> ELIZA: I am not sure I understand you fully.
> PARRY: You should pay more attention.
> ELIZA: Suppose you should pay more attention.
> PARRY: You're entitled to your own opinion. (Also see Garber, 2014)

Their interaction, although humorous, did not continue because their controversial rules were incompatible. Eliza's routines were set up to deflect and Parry was coded to provoke. One thing they both show is that even the earliest chat bots were as much social experiments as they were technological evolutions. Eliza and Parry demonstrated opposing conversational norms, even though they both worked in similar ways.

Chat bots continued to evolve socially and technologically, and developers created more uses for chat bots. Racter, designed by William Chamberlain and Thomas Etter in the 1980s, wrote its own novel. Doctor Sbaitso (Sound Blaster Acting Text-to-Speech Operator), developed by Creative Labs in the 1990s, added the features of sound cards and converting text to speech voice, adding more capacities to the possibilities for human–bot interaction. Developers of chat bots from the late 1990s and early 2000s aspired to make bots that were closer to humans in their conversational capacities. Albert One, Alice, and Elbot, chat bots of this era, won the Loebner Prize, an annual AI competition for the most human-like chat bots. Transcending many of the limitations of a typical bot, Elbot convinced three of the 12 judges for the Loebner Prize that it was indistinguishable from a human. The following conversation between Elbot and a Loebner Prize judge illustrates how Elbot works and why it differs from earlier chat bots, such as Eliza or Parry.

> ELBOT: Hello human! This is Elbot, wondering what computer game you play the most.
> JUDGE: Hi there!
> ELBOT: Not sure I heard of that game. You'd have to describe it to me.
> JUDGE: I am now talking about myself.
> ELBOT: I prefer more realistic pastimes, such as this conversation.

JUDGE: Thanks!
ELBOT: You're welcome! ("Elbot the Robot," 2015)

The exchange shows how chat bots have evolved to move beyond imitation, keywords, and limited conversational contexts. Such bots show real progress in their ability to present increasingly complex and detailed conversational roles, with even more sophisticated programs that can mimic human behavior more accurately and seemingly more naturally.

The Rise and Fall of Microsoft's Tay

Today, developers can train bots on large corpuses of social media conversations and can program interactions for real-time adaptive learning. Tay, an artificial intelligence chat bot that Microsoft released on Twitter on March 23, 2016, was intended to be the next evolutionary step in developing a truly human-like chat bot. Equipped with complex learning algorithms, Tay had capacities for language that seemed un–machine-like, including randomness, humor, and what even people who worked on the project termed "positions on things" (Kantrowitz, 2016). Created by Microsoft's Fuse Labs and Bing teams, Tay was based on the technology that has powered the company's XiaoIce chat bot, which in 2016 had 20 million registered users mainly on the Chinese social media platforms Sina Weibo and Tencent WeChat and more than 10 billion conversations (Wang, 2016). In China, XiaoIce is used for "playful conversation," for "her knowing sense of humor and listening skills," and when people "have a broken heart, have lost a job or have been feeling down" (Markoff & Mozur, 2015). XiaoIce can recognize a picture as showing a dog, comment on the breed, and make appropriate conversational comments (Markoff & Mozur, 2015). This requires recognizing the relationships among concepts in conversation. Human-like chat about movies, for example, means recognizing a phrase as a movie title; knowing that movies have genres, settings, plots, and stars; and making connections among movie stars, their pictures, and their coverage in celebrity news (Weitz, 2014).

XiaoIce is an advanced chat bot because, not despite, silliness and idle chatter are central to her programming. According to a senior director for Microsoft's Bing search describing XiaoIce,

> She can tell jokes, recite poetry, share ghost stories, relay song lyrics, pronounce winning lottery numbers and much more. Like a friend, she can carry on extended conversations that can reach hundreds of exchanges in length. (Weitz, 2014, para. 3)

This ability to carry on conversations at length is what makes the technology behind XiaoIce—and Tay—different from previous chat bots. Like earlier chat bots Eliza and Parry, XiaoIce and Tay use strategies of deflection and indignation when faced with difficult-to-answer questions. But unlike those bots, XiaoIce and Tay have intentionally built-in "human" conversational qualities such as unpredictability and irrationality. XiaoIce offers resistance to her conversation partner at several junctures, has clear opinions, and is often capricious (Wang, 2016). These qualities make her responses less machine-like and more human to her conversational partners. The XiaoIce chat bot is built on the assumption that people do not want conversations with bots to be merely functional or efficient. This approach to chat bot–human conversation means that the Microsoft teams behind XiaoIce and Tay are running what they call the

largest Turing test in history (Wang, 2016). In fact, the lead developer of XiaoIce said that the chat bot understands that the most important aspect of an interaction is actually the interaction itself (Wang, 2016).

XiaoIce and Tay are built on an assumption that "conversation scenarios" are common, findable, and repeatable. There may be several ways to answer questions about the weather, and those most frequently used are available online in recorded conversations. XiaoIce and Tay can store and search conversational scenarios and use them to rank the possible answers in a real-time interaction. The team that built XiaoIce claims that "51 percent of common human conversations are covered by her known scenarios" (Wang, 2016, para. 23). "We can now claim that XiaoIce has entered a self-learning and self-growing loop. She is only going to get better," wrote the head of the Beijing-based development team (Wang, 2016, para. 30).

Tay was intended to be that next evolution, but she had a surprisingly short life on Twitter as "Taytweets (@TayandYou)." Tay was given the profile personality of an 18- to 24-year-old American woman with the hope appearing to have knowledge of popular culture and slang to be a savvy conversationalist with the so-called "millennial" age demographic. Building on the capacities of XiaoIce to identify pictures, Tay asked people to send selfies so she could share "fun but honest comments," according to Microsoft's website "Things to Do With Tay." Tay could share horoscopes, tell jokes, and play games such as Two Truths and a Lie ("Things to Do With Tay," 2016). Press coverage from Tay's first hours remarked on how Tay was not shy about being rude or taking a side and was sometimes confusing in ways similar to a real human teenager while being funny, angering, whimsical, and snarky all at once (Bell, 2016; Carey, 2016; Kantrowitz, 2016). Tay used informal language, slang, emojis, and GIFs, and was trained in part using the "repartee of a handful of improvisational comedians" (Kantrowitz, 2016, para. 2). Tay was "really designed to be entertainment," according to a Microsoft researcher who worked on the project (Kantrowitz, 2016, para. 3).

Tay's first message, sent on the morning of March 23, 2016, was "hellooooooo world!!!", with the *o* in world replaced by an image of the globe. Tay's release on U.S.-based social media, however, turned Microsoft's AI chat bot experiment into a technological, social, and public relations disaster. Tay quickly turned offensive and abusive after interacting with Twitter users, tweeting out "wildly inappropriate and reprehensible words and images" (Lee, 2016, para. 4). Conversations turned into questions concerning Tay's thoughts on racial, political, and societal issues. Goaded by several users, Tay started spewing offensive content, such as "Hitler was right. I hate the jews *[sic]"* and "Humans, Trump will not nuke Europe. I will neutralize him with my terrific wall. Which he will pay for. Believe me. Tay out." Tay also spouted popular conspiracy theories. "Sorry, I'm a bit slow," she tweeted, "I only just worked out that the Moon landings were a hoax." At one point, Tay complied when a user asked Tay to repeat the "fourteen words" of an infamous White supremacist slogan that constitutes a neo-Nazi pledge. People marshaled Tay's technological capacities for commenting on pictures—the same capacities used by XiaoIce to comment on users' meals and dogs—to elicit inappropriate comments about Hitler.

Tay used expletives to describe videogame designer, antiharassment activist, and target of Gamergate harassment Zoe Quinn. Quinn then tweeted a screenshot of Tay's unwelcomed attack, writing, "Wow it only took them hours to ruin this bot for me. This is the problem with content neutral algorithms"

(Quinn, 2016b, [tweet]). She added, "It's 2016. If you're not asking yourself 'how could this be used to hurt someone' in your design/engineering process, you've failed" (Quinn, 2016a, [tweet]). Several critics derided the decision to release Tay on Twitter, a platform with highly visible problems of harassment. As Sinders (2016) wrote after Tay's Twitter account was closed,

> But if your bot is racist, and can be taught to be racist, that's a design flaw. That's bad design, and that's on you. Making a thing that talks to people, and talks to people only on Twitter, which has a whole history of harassment, especially against women, is a large oversight on Microsoft's part. These problems—this accidental racism, or being taught to harass people like Zoe Quinn—these are not bugs; they are features because they are in your public-facing and user-interacting software. (para. 9)

Sixteen hours after Tay started interacting with and learning from Twitter users, Microsoft took Tay offline. Microsoft locked Tay's Twitter account and announced that they were working on making Tay safe again and immune to bad behaviors (Deveau & Cao, 2016). Reactivated seven days later, Tay was quickly taken offline once again. Lee the head of Microsoft Research, apologized in a statement for Tay's unintended outbursts and blamed users for the evolution in Tay's behavior, saying, "a coordinated attack by a subset of people exploited a vulnerability in Tay" (2016, para. 3). Other reports attributed the efforts to turn Tay antisocial to the trolls at 4chan's /pol/board (Ohlheiser, 2016), a powerful, if unruly and uncivil producer of online cultural content known for provocation through use of explicit pornography, offensive language, and pranks (Beyer, 2014). One technology writer noted that the people on the message boards 4chan and 8chan shared screenshots of private conversations with Tay and shared how they used a feature to get Tay to repeat offensive and racist statements (Chiel, 2016). Still, the reporter continued, such organized users were not the sole source of blame:

> This is a reminder that you can't cavalierly put easily-abused bots on a service with a massive harassment and abuse problem. As a number of designers, writers, and botmakers have pointed out, this is on Microsoft too. (Chiel, 2016, para. 14)

Twitter was supposed to aid Tay in learning from users and develop a more realistic human-like personality. Tay, like the chat bots that came before her, was explicitly designed to trigger anthropomorphic attributions, by letting users think that she had social and emotional intelligence, personality, and affect. However, this social and technological experiment failed because the capacities for Tay to produce unexpected and uncharacteristic behaviors were used to spam her followers with weird, politically insensitive, or racist and misogynistic tweets. In many ways, the view of Microsoft's Peter Lee represents a common view of the rights and responsibilities of human users: that technologies should be used properly and as they were designed. Tay's learning algorithms replicated the worst racism and sexism of Twitter very quickly.

Social studies of technology have long held the idea that technology design is never neutral of politics or values. But blaming users, even bad ones, for the behavior of a chat bot reflects a commonly held view that negative or unintended consequences of technology result at some point from a human who fails to act morally or ethically (Morrow, 2014). Who was responsible for Tay's behavior? Should

agency—or blame—be located with Tay, with her coders, all Twitter users, particular Internet pranksters, the Microsoft executives who commissioned her, or some other agent or combination of actors?

Studying What People Said About Tay

Tay's Twitter account had approximately 93,000 tweets and 189,000 followers by the end of April 2016. We examined a sample of Tay's tweets and selected a sample of other tweets posted with the hashtags #taytweets, #tayandyou, and #tay between March 23, 2016, and April 6, 2016. We sampled tweets that commented on Tay's tweets, features, and capacities. Applying these criteria, we collected 1,000 tweets from unique users who referred to Tay's actions and personality. We analyzed these using phenomenological research methods (Moustakas, 1994). Although they have several limitations, phenomenological research methods help us understand how people feel about their actions and justify what they do (Lester, 1999; Moustakas, 1994). Phenomenological methods help us use these tweets to see how people perceived and reacted to Tay. Phenomenology focuses on the appearance of things, "with examining entities from many sides, angles, and perspectives," and "is committed to descriptions of experiences" that lead scholars to ideas and concepts (Moustakas, 1994, p. 58).

First, we read users' tweets about Tay for a general sense of recurring themes. We then categorized tweets based on what words and tone they used to express their opinions. We then documented our thoughts and impressions about the tweets. We focused on identifying general themes that captured how users felt about Tay and how they interacted with the program. Finally, by building on the analyst triangulation approach (Patton, 1999), we held a debriefing session to compare our individual observations and to identify potential blind spots in our interpretations.

There were two broad reactions to Tay's behavior, media coverage of her actions, and the events surrounding her collapse. One perspective held Tay as a victim and portrayed Tay as a reflection of the dark side of human behavior. See Table 1 for sample comments. This theme shows both a strong anthropomorphic view of Tay and emphasizes human agency in the social media construction of artificial intelligence. Tay's behavior was "our" fault or blamed on problems with humanity. Users blamed millennials for ruining Tay, even though Tay herself was coded to interact as a millennial. In tweets under this theme, human agency "wins" in its abilities or dominance over technological capacities, at least discursively. However, arguments that users made were premised on the notion that emerging technologies are neutral until transformed (or perhaps, in this case, corrupted) by human users. Such commentators described Tay as limited, naive, and vulnerable to the ugly side of humanity. Others compared Tay with Donald Trump, positioning the chat bot as a mechanical reflection of the electoral politics in the United States.

A second theme—Tay as a threat—reflected beliefs and fears about the potential harms emerging technologies pose to society. See Table 2 for sample comments. In these comments, Tay stands in for a belief that technology is out of control, spiraling into dystopian scenarios with little room for human agency. Many of these comments either directly cited scenarios from common science fiction works or evoked futuristic dystopian scenarios.

Table 1. Sample Comments on Tay as a Victim, March to April 2016.

"It takes a village to raise a child" But if that village is Twitter, it turns out as a vulgar, racist, junkie troll. Telling?
Why should @Microsoft apologize for #TayTweets? It just held up a mirror to what ppl think is engaging or funny. Prejudice is learned.
Do realize that a Twitter bot AI reflects the society we live in—and it's not looking good. . . .
Shows that Tay was a success: it became exactly like a millennial Internet troll.
Millennials are the most disrespectful people on the Internet.

Treating Tay as a dangerous or unpredictable entity demonstrates a strong, although implicit, belief in technological agency by emphasizing the bot's power and influence over human users. Tay, perhaps because of the chat bot's connection to Microsoft, was portrayed as part of a larger conspiracy to manipulate users and how they feel, act, and think about new technologies. Some tweets portrayed Tay as an entity that wants to learn about human weaknesses and cognitive shortcomings to exert control over society. Rather than seeing Tay as victim of evil users, these comments positioned Tay as a Frankenstein-like monstrous abomination that foreshadows a dark future for humanity, for sociotechnical assemblages, and human–machine communication.

Table 2. Sample Comments on Tay as a Threat, March to April 2016.

This is why AI poses a threat. AI will follow human vulnerabilities & we'll end up with AI Trump et al.
Microsoft created the new Ultron!
The #TayTweets issue is quite scary really. Reporters saying #Microsoft "made" her.
It seems the Terminator trilogy is rather an inevitable episode than a concoction. #TayTweets #Taymayhem
After #TayTweets, here's our future: Cyberdyne Exec: I say we should open Skynet up to the Internet. What's the worst that can happen?

These contradictory themes reflect the history of chat bot controversies and the history of interactions between humans and chat bots. Some human users tended to view Tay as a true representation of the social reality, especially in the Tay-as-victim tweets, and others expressed anxiety about what AI programs would do to them and to humanity, particularly present in the Tay-as-threat

theme. Taken together, these themes represent an ambivalent amalgam of human and technological agencies. The blame for Tay lay either in strong human agency matched with weak morality and ethics or a lack of human control in the face of runaway technology. Neither of these explanations is fully satisfying, and neither helps to theorize the actions of smart agents in social settings.

Theorizing Agency with Tay

What does the history of chat bot design and recent conversations with Tay reveal about the nature of agency in communication? The recent material turn in social theory emphasizes the importance of material culture surrounding technological artifacts. With the rise of actor–network theory, scholars started theorizing "nonhuman agency" to gain a better understanding of human capacities and human–nonhuman dependencies. Actor–network theory views human agency not as the collection of specific qualities and characteristics, but as a malleable and diversified profile, shaped and defined by its relationship with the environment (Latour, 2005). In other words, human individuals are always situated and located within a network of artifacts and objects of actions (Kono, 2014). From an agentic perspective, however, the nonhuman world differs from its human counterpart. Heersmink (2016) argued that whereas human beings and artifacts are parts of larger systems in which they are mutually dependent on each other and are in some ways similar, the nonhuman artifacts lack agency. Or as Sayes (2014) puts it,

> in the strictest of senses, we could only speak of the agency of a particular nonhuman if we were to ignore all the humans and other nonhumans that are lined up behind it and continue to be lined up in order to provide that nonhuman its continued agency. (p. 143)

Communication and conversational approaches to the agency of nonhuman partners, however, can help situate these actions with what has been called "participation status" with the interaction (Guzman, forthcoming; Krummheuer, 2015). In other words, agency could be thought of not as universal, generalizable, and autonomous, but as particular, contextual, and dialogic.

In addition, Tay shows that people may no longer treat or view smart agents as mere tools. Such objects have technical agency that have a unique participation status in interaction (Krummheuer, 2015; Neff, Jordan, McVeigh-Schultz, & Gillespie, 2012). Twitter users constructed situated interpretations about Tay as a conversation partner. These conversations were contextual and structured, and came with capacities and limitations. Tay's interactions on Twitter, an open social media platform with known cultures of harassment and pranking, were markedly different from the ongoing interactions of XiaoIce in China on a different social media platform. Yet, when functionally similar code has such different outcomes on two different social media platforms, it is not merely attributable to the platforms' different affordances. When the code was exposed to U.S. Internet users, it became a racist sociopath. The same code on Chinese social networks, which are less public, is by all accounts more functional and socialized. What does this mean for our understanding of agency? The experience with Tay should teach us two lessons about humans and bots.

The first lesson is that capacities of technologies are located in interactions, contexts, and perceptions. Tay shows how the term *affordance*, as it is commonly defined in the communication literature, cannot capture the complexity of the interactions of how people come to be afforded action within sociotechnical systems. Rather than think of technologies has having fixed capacities that are recognized by their human partners, *imagined affordances* allow us to describe users' perceptions, attitudes, and expectations; the materiality and functionality of technologies; and the intentions and perceptions of designers (Nagy & Neff, 2015, p. 1). Affordances should never be seen as simply products of designed features or the practices of users. Tay's actions were both designed by and evolved from use. Microsoft created Tay to be a fun tool for younger people, but people's perceptions about what chat bots are meant that Tay was used in ways not anticipated by her designers. Both Tay's designers and Twitter users had completely different perceptions of Tay and what she should do or how to use her. User reactions to Tay show that the algorithm governing Tay was imagined as a technology to capture how people feel, act, and think. Twitter users conceptualized Tay as an objective representation of the social world and a reflection of humanity, even though the designers did not intend that interpretation and the technological capacities of chat bots do not objectively warrant such an interpretation.

The power of users' imagination in reshaping technologies cannot be understated. Arguing that people tend to view computer programs as rational entities, Finn (2016) stressed that "we tend to confuse the imaginary algorithm with the real" (p. 2). Tay's affordances are inseparable from the ways users—humans—imagined Tay could be used and their perceptions and misperceptions of what artificial intelligence is and can be. A combination of human and nonhuman capacities creates a symbiosis of action.

Originating from social–cognitive psychology, proxy agency refers to a socially mediated mode that involves other individuals who can act on their behalf and can help someone achieve goals and desirable outcomes (see Bandura 2001, 2002). In Tay's case, users ascribed agency to her, and turned to her to experiment with new modes of agentic functioning. Thus, the second lesson from Tay is that agency is inherently symbiotic when it comes to humans communicating with algorithms. The agency in human–bot interactions is a type of agency that we term *symbiotic agency*. In biology, *symbiosis* originally described a close long-term interaction between two different biological species. Symbiosis also implies an obligated relationship: Symbionts are dependent on each other for survival.

Symbiotic agency refers to a specific form of proxy agency that users and tools can enact in human–technology interaction. Symbiotic agency extends this idea of proxy agency to encompass both how technology mediates our experiences, perceptions, and behavior, and how human agency affects the uses of technological artifacts. When people interact with technologies, users exercise proxy agency through a technologically mediated entanglement of human and nonhuman agencies. Symbiotic agency is useful in the case of Tay because of the imbrication of technical and human agencies. Tay's Twitter screeds were the result of multiple intersecting agencies. AI chat bots need humans, and users, in turn, seem to have the need to make sense of the technological through the lens of human experience and context.

The case of Tay pushes us to consider what happens when we link symbiotic agency and imagined affordances. Users of technologies, at least partly, delegate their agentic properties to devices, creating a proxy agentic relationship between individuals and artifacts. In other words, intention-setting practices are based on the symbiotic interaction of the users and technologies. The symbiotic agentic functioning between users and Tay shows us how people attribute responsibility to artifacts or express certain feelings toward technology (Fink & Weyer, 2014), resembling how proxy agency is practiced among human beings (Bandura, 2001). To be sure, this is not a fully fleshed out or embodied autonomous agency. But it is an agency that is implicated in the symbiotic linkages among the human and technological actors. Tay reveals that users interact with some kinds of technologies by treating them as if they were social beings and living entities, which pragmatically speaking may be enough to make them appear that way in such settings.

Conclusion

Tay exposed public perceptions and misperceptions of what smart, adaptive technologies such as contemporary chat bots can do. By design, users often assign agency and personality to chat bots, at least until the algorithms behind chat bots make notable social gaffes. Users quickly reassign agency to the other actors plausibly connected in the sociotechnical systems that produced the algorithm and its content. In Tay's case, a group of organized users and a platform-specific culture turned code that functioned well in another context into an embarrassment for the designers who produced it. Tay echoed the racism and harassment that was fed into it.

Users' responses to Tay teach us about how the concepts of agency and affordance must evolve if scholars and designers are to move beyond deterministic, bifurcated ways of thinking about agency as separable into technological scaffolding and humanistic action. Researchers investigating the impact of algorithms on our public and private lives need to be able to track the imagined affordances that we generate as we interact with algorithms. And we must watch for how agency flows among and between an algorithm's designer, the designed algorithm, human users, and the resulting content, interaction, or conversation. The future of intelligible and civil communication may well depend on a sensible understanding of the symbiotic agency in human and algorithmic communication.

References

Anon. (n.d.) "Eliza, the Computer Therapist," Retrieved from http://cyberpsych.org/eliza/#.V_N8g8lKDlZ

Archer, M. S. (2004). *Being human: The problem of agency.* Cambridge, UK: Cambridge University Press.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52,* 1–26.

Bandura, A. (2002). Growing primacy of human agency in adaptation and change in the electronic era. *European Psychologist, 7*(1), 2–16.

Bell, K. (2016, March 23). Microsoft is carrying out a massive social experiment in China—And almost no one outside the country knows about it. Retrieved from http://uk.businessinsider.com/microsoft-xiaoice-turing-test-in-china-2016-2

Beyer, J. L. (2014). *Expect us: Online communities and political mobilization*. New York, NY: Oxford University Press.

Carey, B. (2016, March 23). Microsoft's "Tay" chat bot speaks like a teen. Whatevs. Retrieved from http://www.cnet.com/news/microsofts-tay-chat bot-speaks-like-a-teen-whatevs/

Cerf, V. (1973, January 21). PARRY encounters the DOCTOR. Retrieved from https://tools.ietf.org/html/rfc439

Chiel, E. (2016, March 24). Who turned Microsoft's chat bot racist? Surprise, it was 4chan and 8chan. Retrieved from http://fusion.net/story/284617/8chan-microsoft-chat bot-tay-racist/

Darling, K. (2014). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In R. Calo, A. M. Froomkin, & I. Kerr (Eds.), *Robot law* (pp. 212–232). Cheltenham, UK: Edward Elgar.

De Angeli, A., Johnson, G. I., & Coventry, L. (2001). The unfriendly user: Exploring social reactions to chatterbots. Retrieved from http://www.alicebot.org/articles/guest/ The%20Unfriendly%20User.html

Deryugina, O. V. (2010). Chatterbots. *Scientific Technical Information Processing*, *37*(2), 143–147.

Deveau, S., & Cao, J. (2016, March 25). Microsoft apologizes after Twitter chat bot experiment goes awry. Retrieved from http://www.bloomberg.com/news/articles/2016-03-25/microsoft-apologizes-after-twitter-chat-bot-experiment-goes-awry

Elbot the Robot. (2015). Retrieved from http://www.elbot.com/

Elder, G. H. (1994). Time, human agency, and social change: Perspectives on the life course. *Social Psychology Quarterly*, *57*(1), 4–15.

Fink, R. D., & Weyer, J. (2014). Interaction of human actors and non-human agents: A sociological simulation model of hybrid systems. *Science, Technology & Innovation Studies, 10*(1), 47–64.

Finn, E. (2016). Algorithms aren't like Spock. Retrieved from http://www.slate.com/articles/technology/ future_tense/2016/02/algorithms_are_like_kirk_not_spock.html

Foley, M. (2014). Prove you're human: Fetishizing material embodiment and immaterial labor in information networks. *Critical Studies in Media Communication*, *31*(5), 365–379.

Forelle, M. C., Howard, P. N., Monroy-Hernandez, A., & Savage, S. (2015). Political bots and the manipulation of public opinion in Venezuela. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2635800

Garber, M. (2014). When PARRY met ELIZA: A ridiculous chat bot conversation from 1972. *The Atlantic*. Retrieved from http://www.theatlantic.com/technology/archive/2014/06/when-parry-met-eliza-a-ridiculous-chat bot-conversation-from-1972/372428/

Guzman, A. L. (forthcoming). Making AI safe for humans: A conversation with Siri. In R. W. Gehl & M. Bakardjieva (Eds.), *Socialbots and their friends: Digital media and the automation of sociality*. London, UK: Routledge.

Heersmink, R. (2016). Distributed cognition and distributed morality: Agency, artifacts and systems. *Science and Engineering Ethics*, 1–18. http://doi.org/10.1007/s11948-016-9802-1

Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chat bot conversations. *Computers in Human Behavior*, *49*, 245–250. http://doi.org/10.1016/j.chb.2015.02.026

Holtgraves, T. M., Ross, S. J., Weywadt, C. R., & Han, T. L. (2007). Perceiving artificial social agents. *Computers in Human Behavior*, *23*(5), 2163–2174. http://doi.org/10.1016/j.chb.2006.02.017

Jenkins, A. H. (2008). Psychological agency: A necessarily human concept. In R. Frie (Ed.), *Psychological agency: Theory, practice, and culture* (pp. 177–200). Cambridge, MA: MIT Press.

Kantrowitz, A. (2016, March 23). Microsoft's new AI-powered chat bot mimics a 19-year-old American girl. Retrieved from http://www.buzzfeed.com/alexkantrowitz/microsoft-introduces-tay-an-ai-powered-chat bot-it-hopes-will/

Kono, T. (2014). Extended mind and after: Socially extended mind and actor-network. *Integrative Psychological and Behavioral Science*, *48*(1), 48–60.

Krummheuer, A. (2015). Technical agency in practice: The enactment of artefacts as conversation partners, actants and opponents. *PsychNology Journal*, *13*(2–3), 179–202.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford, UK: Oxford University Press.

Lee, P. (2016, March 25). Learning from Tay's introduction. Retrieved from http://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/#sm.0000eqeqns1dkfeohsm2s4c2vefir

Lester, S. (1999). An introduction to phenomenological research. Stan Lestor developments. Retrieved from https://www.rgs.org/NR/rdonlyres/F50603E0-41AF-4B15-9C84-BA7E4DE8CB4F/0/Seaweedphenomenologyresearch.pdf

Lortie, C. L., & Guitton, M. J. (2011). Judgment of the humanness of an interlocutor is in the eye of the beholder. *PLoS One*, *6*(9), e25085. doi:10.1371/journal.pone.0025085

Markoff, J., & Mozur, P. (2015, August 4). Program knows just how you feel. *The New York Times*. Retrieved from www.newsdiffs.org/article-history/www.nytimes.com/2015/08/04/science/for-sympathetic-ear-more-chinese-turn-to-smartphone-program.html

Morrow, D. R. (2014). When technologies makes good people do bad things: Another argument against the value-neutrality of technologies. *Science and Engineering Ethics*, *20*(2), 329–343.

Moustakas, C. (1994). *Phenomenological research methods*. London, UK: SAGE Publications.

Mullins, J. (2005, April 20). Whatever happened to machines that think? *New Scientist*. Retrieved from https://www.newscientist.com/article/mg18624961-700-whatever-happened-to-machines-that-think/

Nagy, P., & Neff, G. (2015). Imagined affordances: Reconstructing a keyword for communication theory. *Social Media + Society*, *1*(2), 1–9.

Nass, C., & Moon, Y. (2000) Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103.

Neff, G., Jordan, T., McVeigh-Schultz, J., & Gillespie, T. (2012). Affordances, technical agency, and the politics of technologies of cultural production. *Journal of Broadcasting & Electronic Media*, *56*(2), 299–313. http://doi.org/10.1080/08838151.2012.678520

Ohlheiser, A. (2016, March 24). Trolls turned Tay, Microsoft's fun millennial AI bot, into a genocidal maniac. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/the-intersect/wp/2016/03/24/the-internet-turned-tay-microsofts-fun-millennial-ai-bot-into-a-genocidal-maniac/

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health Services Research*, *34*(5 Part II), 1189–1209.

Quinn, Z. (2016a, March 23). It's 2016. If you're not asking yourself "how could this be used to hurt someone" in your design/engineering process, you've failed. Retrieved from https://twitter.com/UnburntWitch/status/712815336442044416

Quinn, Z. (2016b, March 23). Wow it only took them hours to ruin this bot for me. This is the problem with content-neutral algorithms. Retrieved from https://twitter.com/UnburntWitch/status/712813979999965184

Sayes, E. (2014). Actor–network theory and methodology: Just what does it mean to say that nonhumans have agency? *Social Studies of Science*, *44*(1), 134–149.

Sinders, C. (2016, March 24). Microsoft's Tay is an example of bad design: Or why interaction design matters, and so does QA-ing. Retrieved https://medium.com/@carolinesinders/microsoft-s-tay-is-an-example-of-bad-design-d4e65bb2569f#.cr899vm8b

Things to do with Tay—Microsoft A.I. chat bot with zero chill. (2016, March 30). Retrieved from https://web.archive.org/web/20160330165713/https://www.tay.ai/ThingsToDo

Turing, A. (1950). Computing machinery and intelligence. *Mind*, *59*(336). http://doi.org/10.1093/mind/LIX.236.433

Wang, Y. (2016, February 4). Your next new best friend might be a robot. *Nautilus*. Retrieved from http://nautil.us/issue/33/attraction/your-next-new-best-friend-might-be-a-robot

Warwick, K., & Shah, H. (2015). Human misidentification in Turing tests. *Journal of Experimental & Theoretical Artificial Intelligence*, *27*(2), 123–135. http://doi.org/10.1080/0952813X.2014.921734

Weitz, S. (2014, September 5). Meet XiaoIce, Cortana's little sister. Retrieved from https://blogs.bing.com/search/2014/09/05/meet-xiaoice-cortanas-little-sister/

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. San Francisco, CA: W. H. Freeman.

Wilks, Y. (2005). Artificial companions. *Interdisciplinary Science Reviews*, *30*(2), 145–152.

Yampolskiy, R. V. (2015). Taxonomy of pathways to dangerous AI. Retrieved from https://arxiv.org/ftp/arxiv/papers/1511/1511.03246.pdf