An Empirical Analysis of Human-Bot Interaction on Reddit

Ming-Cheng Ma, John P. Lalor
Department of IT, Analytics, and Operations
University of Notre Dame
`mma4@alumni.nd.edu`, `john.lalor@nd.edu`

Abstract

Automated agents ("bots") have emerged as an ubiquitous and influential presence on social media. Bots engage on social media platforms by posting content and replying to other users on the platform. In this work we conduct an empirical analysis of the activity of a single bot on Reddit. Our goal is to determine whether bot activity (in the form of posted comments on the website) has an effect on how humans engage on Reddit. We find that (1) the sentiment of a bot comment has a significant, positive effect on the subsequent human reply, and (2) human Reddit users modify their comment behaviors to overlap with the text of the bot, similar to how humans modify their text to mimic other humans in conversation. Understanding human-bot interactions on social media with relatively simple bots is important for preparing for more advanced bots in the future.

1 Introduction

People across the world engage with each other on social media sites for personal, professional, and entertainment-related reasons. In recent years automated agents ("bots") have become more prevalent on social media (Ferrara et al., 2016). Bots engage with human users on social media platforms via the platforms' application programming interfaces (APIs). By listening to content using the API, bots are coded to engage based on specific keywords or phrases that are used by human users.

As bots become more prevalent on social media, more and more humans find themselves engaging with these bots. These humans may or may not be aware of the fact that the bots are not humans. There is a need to study how humans and bots interact with each other, and to analyze how bots influence the way humans engage on the platforms. On certain platforms bots are opaque (that is, you do not know if a user is a bot). For example, on

Twitter most bots are opaque, and there is a steady stream of research on detecting bots and analyzing the effect they have on the behavior of human Twitter users e.g., (Wang, 2010; Chu et al., 2012; Clark et al., 2016). However, on Reddit many bots are open and explicit about their botness. Reddit users engage with each other on topic-focused communities ("subreddits"). If users are knowingly interacting with a bot, does the bot influence what the users will comment? More specifically, our research question for this work is: *How does engagement with a known bot influence human behavior on social media?*

In this work we present a case study of a bot that engages frequently on a small number of subreddits to do a deep dive into the way the bot interacts with humans on the subreddits. The bot's comments are pre-defined and randomly selected, so there is no true "interaction" per-se with respect to how the bot replies to human users. We are interested in seeing if humans reply to this simple bot in interesting ways. As bots become more advanced with improving NLP technologies, the effect on human-bot interaction will become more pronounced. Therefore, it is important to understand interaction dynamics with simple bots to theorize about how these interactions may change with more advanced bots.

As a result of our analyses we identify two interesting findings. First, we find that the sentiment of the (randomly generated) bot comment has a significant effect on the sentiment of the human reply comment. Second, we compare the text content of the bot comments and human replies and find evidence of *lexical entrainment*, where humans overlap with bots in terms of their text comments, consistent with known patterns of conversation between humans (Beňuš et al., 2014).

Variable	M1	M2	M3
Parent Sentiment	-0.008 (0.011)	0.077 (0.012)***	-0.008 (0.011)
Bot Sentiment	0.039 (0.009)***	0.084 (0.009)***	0.039 (0.008)***
Parent Child Same User	0.036 (0.006)***	-	-
Parent Sentiment x Bot Sentiment	-0.022 (0.023)	-0.042 (0.023)*	-0.022 (0.022)
Parent Sentiment x PC Same	0.085 (0.016)***	-	-
Bot Sentiment x PC Same	0.046 (0.013)***	-	-
PS x BS x PC Same	-0.020 (0.032)	-	-

Table 1: Regression results. Standard errors in parentheses. $*p < 0.1$, $***p < 0.001$

2 Data Collection

In this work we focus on interactions between humans and a single-purpose Reddit entertainment bot. We analyze interactions between Reddit users and bobby-b-bot, a Reddit bot inspired by the Game of Thrones books and TV series.[1] Bobby-b-bot posts are randomly selected quotes from a Game of Thrones character, Robert Baratheon.

We selected the bobby-b-bot for our analyses because the bot is a purely "entertainment" bot, in contrast to many Reddit bots that perform some utility (e.g., text summarization or subreddit moderation). The bot source code is open-sourced and available, so we can see how the bot identifies comments on Reddit that it should reply to. To activate bobby-b-bot, a Reddit user posting on one of the specified subreddits[2] must include some variation of the bot's name in their comment. Once activated, the bot will reply to the comment with a randomly selected quote from the GoT books.

To collect all bobby-b-bot comments from Reddit, we pulled the bot comment data as comment triples from the Pushshift Reddit API[3]. In each triple, there is a human post ("parent"), followed by the bobby-b-bot reply, and finally another human post ("child"). We extracted 126,329 bot comments, spanning from 2017/10/23 GMT-4 to 2020/06/14 GMT-4. There are 95,206 (75%) positive parent comments and 31,123 (25%) negative parent comments. For bot comments, there are 75,634 (55%) positive bot comments and 50,695 (45%) comments. When accounting for those bots comments where another user replied, we were left with 16,124 parent-bot-child comment triples. Among these child comments, we identified 12,109 (75%) positive comments and 4,015 (25%) negative comments.

3 Sentiment Analysis

We first analyze the Reddit comment data using sentiment analysis (Liu et al., 2010). In particular, our goal is to determine whether the sentiment of a bot's comment has an effect on the sentiment of comments made in reply. We use the VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon as our sentiment tool (Hutto and Gilbert, 2014). VADER is a specific tool that is designed for analyzing social media texts. It can generate the sentiment score based on the unlabeled given texts showing the polarity of sentiment (how positive or negative). We fit a linear regression model to predict the VADER sentiment score of the child comments in our data set. Our three independent variables were: the parent sentiment score, the bot sentiment score, and an indicator variable for whether the parent and child in the triple are the same user. We included all interactions between independent variables in our model.

The regression results provide some interesting observations (Table 1, M1). Parent sentiment does not have a significant effect on the child sentiment, whereas bot sentiment has a significant positive effect on child sentiment. The indicator variable has a significant positive effect on child sentiment as well. The interaction between parent sentiment and the indicator variable is significant. Because the parent sentiment is by itself not significant, there is potentially a crossover interaction. Therefore, we ran two additional regression models, splitting the data based on whether the parent and child were the same user (M2) or different users (M3).

When the parent and child users are the same, both parent and bot sentiment have a significant positive effect on the child comment sentiment (Table 1, M2). That the parent comment sentiment affects child comment sentiment is intuitive, as the same user will be more likely to be consistent in terms of sentiment in the conversation with the bot. However, it is interesting that the bot sentiment has

(a) Parent-child entrainment.

(b) Bot-child entrainment.

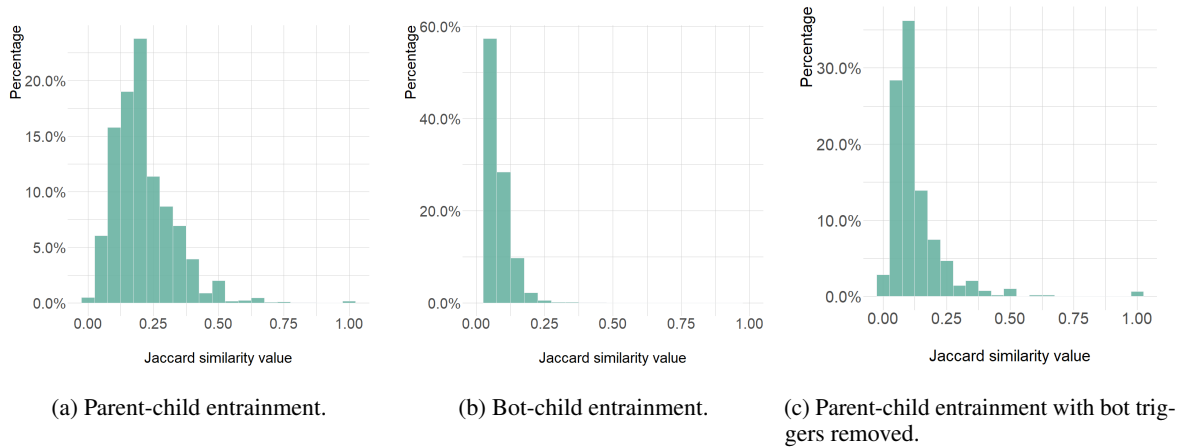(c) Parent-child entrainment with bot triggers removed.

Figure 1: Histograms of Jaccard scores for the three comment pairs in our analyses.

a positive affect as well. When the parent and child users are different, there is no longer a significant effect from the parent comment sentiment, but the bot comment sentiment is still significant (Table 1, M3). That a randomly generated quote can affect the sentiment of a comment written by a human is interesting. This result is consistent with prior work on bots influencing human behavior, but here there is no incentive or end goal for which the bot was built (Nass et al., 1995; Bell et al., 2003; Coulston et al., 2002). The human influence is only based on the entertainment value.

4 Textual Overlap

We next consider how Reddit users modify the text of their comments in response to the bot (*lexical entrainment*). Lexical entrainment is when speakers use words that overlap with words of their conversational partner (Beňuš et al., 2014). Prior work has shown that humans will adapt their speech when interacting with computers (Nass et al., 1995; Bell et al., 2003; Coulston et al., 2002). Computer agents have been designed to encourage individuals to adapt their speaking rate or amplitude (Bell et al., 2003; Coulston et al., 2002). However, here the bot presents randomly selected text, and the only aim of the bot, insofar as it has an aim, is to entertain. Will humans adjust their comments when interacting with the bot?

For our entrainment metric we use Jaccard similarity (Levandowsky and Winter, 1971). Jaccard similarity is a ratio of overlapping tokens to total tokens between two comments, and from 0 (no overlap) to 1 (perfect overlap). Entrainment metrics in prior work are token-focused, where entrainment

for specific keywords are measured (Beňuš et al., 2014). Here we are interested in a global measure, and we therefore use Jaccard similarity to that end.

We measure entrainment in our comment triples across three comment pairs: (1) between parent and child comments, (2) between bot and child comments, (3) and between parent and child comments where the trigger phrase has been removed. For cases 1 and 2 we can compare human-human entrainment with bot-human entrainment. However, we include case 3 so that we can remove examples that may artificially inflate our entrainment results. Recall that the bobby-b-bot is activated by a user including the bot name in a comment. Therefore the phrase "bobby b" (or a variation) appears in all parent comments. In certain cases the child comment also includes the phrase, as the child user also wants to summon the bot. To avoid counting these we remove all instances of "bobby b" from the parent and child comments for our case 3 analysis. In all cases we remove stop words before calculation.

In each case we plot a histogram showing, of those triples where there is some overlap, the Jaccard scores of the relevant pair in the triple (Figure 1). For parent-child entrainment there are 10,723 comment pairs with overlap, for bot-child entrainment there are 4,527 pairs, and for parent-child entrainment with the bot triggers removed there are 2,728 pairs. First, the vast majority of overlap between humans (i.e., the parent-child comment pairs) is a result of the child comment user calling the bot again via the "bobby b" keyword. Once this is removed the number of instances with overlap drops significantly. However, we still observe some positive amounts of overlap between humans. That this overlap occurs with another conversant

	Parent comment	Bot Comment	Child comment
1	Bobby B is our true savior, he helped us cope with *season 8*.	CAREFUL, NED! CAREFUL NOW!	That's right, Bobby B, we have to be <u>careful</u> when watching *Season 8* or we might lose braincells.
2	I'm usually always for more but I actually re-watched a clip from e2 the other day (when it's decided Lady will be killed in Nymeria's place) and forgot how disappointed I was in how Bobby B handled it when Ned questioned the order. Been sad about it	SURROUNDED BY LANNISTERS! EVERY TIME I CLOSE MY EYES I SEE THEIR BLONDE HAIR AND THEIR SMUG, SATISFIED FACES!	I suppose you were <u>surrounded</u> at that time and couldn't do much else, I get it. This bot is alive Rip Lady
3	*Bobby B*, c'mon man, you have to admit it's pretty similar to something you say all the time.	GODS WHAT A STUPID <u>NAME</u>!	'What? Ros? Well it's probably short for something like Roselyn, or knowing GRRM, maybe like Rosmallysyn or some moderately real sounding anglo name that isn't actually a name. I'm surprised he went with something as normal as Robert for you. Handy nicknames like *Bobby B* are pretty common though. Like Ned. Ned is a normal <u>name</u>, but I don't think Eddard is a name in real life. Thoughts?
4	*Why do you keep calling me Ned Bobby B?*	A DOTHRAKI HORDE ON AN OPEN FIELD, NED!	*Why do you keep calling me Ned Bobby B?*

Table 2: Examples of comment triples from the data set. Parent-child comment overlaps are *italicized*, and bot-child overlaps are <u>underlined</u>.

(the bot) in between suggests that the human users are conversing between themselves even though the bot is in the middle. This could be due to the fact that the bot randomly generates text, and is not contributing to the conversation per se. We found that some of the human-human entrainment occurs when the parent and child user are the same, suggesting that the individual has the bot as a conversant but is in effect speaking with herself.

That there are positive amounts of overlap between humans is interesting but not surprising. More surprising is the presence of entrainment between the bot and the child comment user. While evidence of human-machine entrainment has been found in prior work (Bell et al., 2003; Brennan, 1996; Coulston et al., 2002), such research involved bots that were programmed with a purpose, so the bots were designed to generate relevant content to interact with humans. However, here the text for the bobby-b-bot is randomly sampled from a list of pre-existing scripts. People still overlap with bot texts to some positive degree even though there is no need for them to modify their speaking behavior in the conversations.

We sampled instances where there is evidence of entrainment to see why human users are interacting with bobby-b-bot (Table 2). In certain cases users will address the bot directly, but even in doing so will entrain with the bot as well, if only in a small way (e.g., a word or two, Table 2 line 1). However, there are times when a user will engage with the bot as part of the flow of conversation (Table 2, lines 2 and 3). Here the user takes the randomly-generated bot response as an actual response and continues the conversation. Finally, in certain cases it seems that users simply want to engage with the bot to see what comments the bot will post, and therefore repeat themselves to re-trigger the bot in

an extended conversation thread (Table 2, line 4).

It is intriguing to see that the child comment adopts words both from the parent comment and the bot comment. Even though the humans are interacting with a bot, and a simple bot at that, they still will try to behave like they are talking to other human beings. Even when the overlaps are just one or two words within the text, those words serve as an important role for human referencing. By mentioning the key common words, humans can still efficiently keep the conversation going.

5 Related Work

Much of the recent work regarding bots has been in the area of bot detection, in particular bots on Twitter (Chu et al., 2012; Clark et al., 2016; Ferrara et al., 2016). On Reddit there are some opaque bots (Hurtado et al., 2019), however in many cases bots are transparent, typically by including the word "bot" in the username. Prior work has looked at human-bot cooperation for subreddit moderation (Jhaver et al., 2019), but to the best of our knowledge this is the first work to study human-bot interaction and comment pattern effects on Reddit. This work also diverges from prior work on human-bot interaction (Bell et al., 2003; Brennan, 1996; Coulston et al., 2002). In most work where the influence of bots is investigated, the bot is purposefully programmed or designed to elicit some response or interaction. However, bobby-b-bot comments are randomly selected snippets of text, and the bot's purpose is to entertain.

6 Conclusion

In this work we present a detailed analysis of the effects of a single social media bot on human communication. While our sample is limited, the num-

ber of such social entertainment bots is only going to increase with more advanced text generation capabilities. The bobby-b bot we studied uses a simple random text selection algorithm to "interact" with users on Reddit. We have shown that even a simple bot can have an effect on the way individuals communicate on the platform. Understanding human-bot interaction when bots are simple is key for building theories of interaction for when the bot technology improves (e.g., a bot that is built on top of GPT-3, Brown et al. 2020).

In this work we consider a single Reddit bot as a case study. Our goal was to determine whether a bot built for entertainment, and not meaningful back and forth with human users, had an effect on human user sentiment and word selection. Certain details of how the bobby-b-bot is implemented (e.g., randomly selecting what text will be posted) are not consistent with the normal flow of conversation. However, we see that in response to the bot's post, humans are matching keywords that the bot used. So in the context of short, discrete interactions on social media, it would be interesting to see if this behavior holds with other bots. An important direction for future work would be to extend this further and conduct similar studies on a wider range of human-bot interactions to see if the results are more broadly applicable. In particular, is this effect more prevalent in goal-driven bots (i.e., bots that seek to change opinion or raise awareness) than in bots that exist for entertainment purposes?

Typically, lexical entrainment is *targeted*, that is, specific keywords are investigated to see if there is overlap (Brandstetter et al., 2017; Iio et al., 2015). However in this work we consider global entrainment, where any overlap (stop words excluded) is tracked. Because the interactions are short, we consider all overlap meaningful. While Jaccard similarity is an appropriate metric in terms of calculating the overall token overlap, more sensitive metrics could also be considered to incorporate weights for different types of tokens (e.g., rare words).

Prior work on human-machine lexical entrainment looked at conversations between humans and physical robots in a shared space, or conversational agents (Hoegen et al., 2019). More generally, the entrainment phenomenon is usually studied over the course of a conversation, to determine if conversation participants imitate each other's conversation styles. However in this work we look at discrete human-bot interactions on social media.

We find evidence of entrainment in short interactions with a technically simple bot. That humans are imitating the bots in these discrete interactions is an interesting result. Future work to investigate this on a larger bot data set is needed to determine if this behavior is wide-spread on Reddit. Even beyond Reddit, bots on other social media platforms such as Twitter may be able to influence human responses without an extended back-and-forth to establish trust.

Acknowledgments

References

Linda Bell, Joakim Gustafson, and Mattias Heldner. 2003. Prosodic Adaptation in Human-Computer Interaction. page 4.

Štefan Beňuš, Agustín Gravano, Rivka Levitan, Sarah Ita Levitan, Laura Willson, and Julia Hirschberg. 2014. Entrainment, dominance and alliance in supreme court hearings. *Knowledge-Based Systems*, 71:3–14.

Jürgen Brandstetter, Clay Beckner, Eduardo Benitez Sandoval, and Christoph Bartneck. 2017. Persistent lexical entrainment in hri. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 63–72.

Susan E Brennan. 1996. Lexical Entrainment in Spontaneous Dialog. page 4.

Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2012. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on Dependable and Secure Computing*, 9(6):811–824.

Eric M Clark, Jake Ryland Williams, Chris A Jones, Richard A Galbraith, Christopher M Danforth, and Peter Sheridan Dodds. 2016. Sifting robotic from organic text: a natural language approach for detecting automation on twitter. *Journal of computational science*, 16:1–7.

Rachel Coulston, Sharon Oviatt, and Courtney Darves. 2002. Amplitude Convergence in Children's Conversational Speech with Animated Personas. page 4.

Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Communications of the ACM*, 59(7):96–104.

Rens Hoegen, Deepali Aneja, Daniel McDuff, and Mary Czerwinski. 2019. An end-to-end conversational style matching agent. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 111–118.

Sofia Hurtado, Poushali Ray, and Radu Marculescu. 2019. Bot detection in reddit political discussion. In *Proceedings of the Fourth International Workshop on Social Sensing*, pages 30–35.

Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

Takamasa Iio, Masahiro Shiomi, Kazuhiko Shinozawa, Katsunori Shimohara, Mitsunori Miki, and Norihiro Hagita. 2015. Lexical entrainment in human robot interaction. *International Journal of Social Robotics*, 7(2):253–263.

Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. 2019. Human-machine collaboration for content regulation: The case of reddit automoderator. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 26(5):1–35.

Michael Levandowsky and David Winter. 1971. Distance between sets. *Nature*, 234(5323):34–35.

Bing Liu et al. 2010. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2(2010):627–666.

Clifford Nass, Youngme Moon, B. J. Fogg, Byron Reeves, and D. Christopher Dryer. 1995. Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43(2):223–239.

Alex Hai Wang. 2010. Don't follow me: Spam detection in twitter. In *2010 international conference on security and cryptography (SECRYPT)*, pages 1–10. IEEE.