# Bots for Research: Minimising the Experimenter Effect

**Senuri Wijenayake**
swijenayake@student.unimelb.edu.au
University of Melbourne
Melbourne, Australia

**Niels van Berkel**
nielsvanberkel@cs.aau.dk
Aalborg University
Aalborg, Denmark

**Jorge Goncalves**
jorge.goncalves@unimelb.edu.au
University of Melbourne
Melbourne, Australia

## ABSTRACT

Experimenter-induced influences can trigger biased responses from research participants. We evaluate how digital bots can be used as an alternative research tool to mitigate these biases, as based on existing literature. We note that the conversational interactivity provided by bots can significantly reduce biased responses and satisficing behaviour, while simultaneously enhancing disclosure and facilitating scalability. Bots can also build rapport with participants and explain tasks at hand as well as a human experimenter, with the added benefit of anonymity. However, bots often follow a predetermined script when conversing and therefore may not be able to handle complex and unstructured conversations, which could frustrate users. Studies also imply that bots with human-like features may induce experimenter effects as similar to humans. We conclude with a discussion on how bots could be designed for optimal utilisation in research.

## KEYWORDS

Experimenter effect; bots; virtual agents; response bias; social desirability; conversational interactivity

## INTRODUCTION

The 'experimenter effect' is a widely observed phenomenon where study participants tend to alter their behaviour and responses in the presence of an experimenter [26]. Participants can behave as such to agree with subtle, often unconscious, cues extracted from experimenters regarding the study's hypothesis [34], in response to characteristics of the experimenter (*e.g.*, race [6, 20], gender [38], social distance [8]), or to please experimenters [34]. Consequently, collected research data may not fully represent reality, potentially resulting in invalid research outcomes. Thus, minimising the 'experimenter effect' has been a critical concern in several research areas such as HCI/HCI4D [8], Social Psychology [25], and Public Health [7], where user-centred data collection methods (*e.g.*, interviews, user studies) are prominently used.

Following an increasing use of 'bots' (interactive programs which can engage in conversations) in a range of domains such as customer support [11], health care [32], and education [39], recent work is exploring the use of bots as a research tool to minimise the 'experimenter effect', by removing the direct interaction between researcher and participant [10, 14, 21, 33]. This paper offers an overview of the existing literature in this regard, discussing benefits and pitfalls of using bots to minimise researcher-induced effects. We present concrete suggestions on how researchers can use bots to reduce bias in responses, encourage full disclosure in sensitive topics, scale participation, while facilitating opportunities for clarification and comprehension to participants. We then discuss the current limitations of this approach, recommending avenues for future work.

## THE EXPERIMENTER EFFECT

Research participants may adjust their behaviour and opinions in the presence of human experimenters, with the intention of contributing towards what they perceive to be the experimenters' preferred study outcome [34]. For example, in a usability study where participants were required to evaluate two mobile video players, participants were in favour of the video player which they thought to be developed by the experimenter, despite being clearly inferior in quality to the alternative [8]. Similarly, in a study by Brown et al. [2], participants who were asked to 'naturally' use a photo sharing mobile application were seen to amplify their app usage with the intention of producing 'useful' research data to assist experimenters. As such, the 'experimenter effect' has significant implications on the validity and reliability of research data collected using user-centred data collection methods.

Moreover, interacting with researchers may also encourage participants to provide socially desirable responses to avoid being perceived negatively, as opposed to providing responses which reflect their true experience or opinion [9]. This is of particular concern when considering sensitive topics such as health [7], sexual behaviour and abuse [38], substance use [5], politics [6] and religion [10], where participants are more inclined to limit their disclosure.

**Experimenter Effect**:
Study participants adjust their behaviour and opinions as a result of an influence exerted by the (assumed) experimenter's expectations or other sociodemographic characteristics.

**Why?**

- To contribute towards desired research outcomes [2, 8, 37].
- For social desirability [6, 7, 9, 10, 12, 38].
- To avoid offending a social group that the researcher may represent (*e.g.,* racial groups) [20].
- To conform with stereotypes about one's social group [7, 23, 36].

**Existing solutions**

- Use well-trained experimenters with sufficient diversity [7].
- Use double-blinded experimental approaches [34].
- Reduce or remove experimenter interaction with participants [7].

Additionally, sociodemographic characteristics of the researcher (*e.g.,* gender, race, social distance) have been shown to amplify biased responses. For instance, men and women provide more accurate reports of their sexual health to same-gender interviewers [38]. Furthermore, female interviewers tend to get higher response rates and disclosure on sexual abuse and mental disorders that might be perceived as victimising [3, 13], whereas reports of substance use are higher in the presence of male interviewers [3].

Race is an equally salient interviewer characteristic. Prior work shows that participants provide honest opinions on racial content to same-race interviewers and alter their presented opinion if it could offend interviewers from a different race [20]. Moreover, an experimenter's race has been seen to trigger stereotype threats, where participants conform to negative stereotypes about one's social group in objective contexts. For instance, studies note that African American participants were more likely to obtain correct answers when facing an African American administrator than in the presence of a White administrator, when answering objective political questions [6] and completing verbal tests [23].

Moreover, previous studies suggest that the perceived social distance between interviewers and participants may impact the probability of respondents reporting substance use [12]. Similarly, Dell et al. [8] state that participants with a similar social status as the experimenters are less likely to provide responses favouring the experimenters' expectations. Thus, the 'experimenter effect' has serious consequences on the quality of the collected data and the validity of subsequent research outcomes.

## Existing Solutions

Literature proposes using well-trained experimenters from diverse backgrounds, and using double-blinded experimental approaches (where both experimenter and participant are unaware of the assigned experimental condition) to minimise experimenter-induced influences [7, 34]. However, recruiting well-trained experimenters with sufficient diversity is expensive and double-blind approaches may not be feasible in a realistic setting.

Moreover, while experimenters do not generally disclose their personal information to participants, certain sociodemographic characteristics like gender and race are difficult to conceal in face-to-face interactions [7]. Even in the absence of visual cues, participants may infer experimenters' characteristics and beliefs through their voice and speech style [7]. As a result, researchers have suggested completely removing experimenter involvement (*e.g.,* self-administered surveys) to minimise the 'experimenter effect' (especially in sensitive research areas) [20]. However, the absence of experimenters have also been seen to cause lower response rates [20], insincere, unreliable and satisficing answers [17], and lower engagement and understanding from respondents [12, 29]. In particular, the ability to converse with

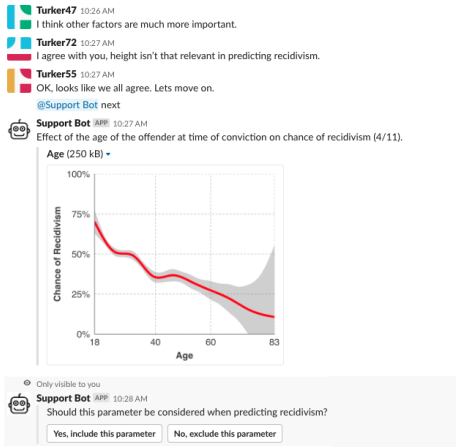


**Figure 1: 'SupportBot', as introduced by Van Berkel et al. [33], presented participants with study instructions, collected participant votes in private, and subsequently initiated public discussions among participants.**

the experimenter (or conversational interactivity) has been seen to facilitate higher comprehension and engagement [12, 29] and conscientious responses [12] among participants.

## BOTS AS A RESEARCH TOOL

We examine the use of bots as a research tool to conduct user-centred experiments with minimal or no involvement from human experimenters. Bots have the ability to facilitate conversational interactivity to study participants, while simultaneously minimising exposure to subtle, unwanted verbal and non-verbal cues provided by human experimenters that could induce desirable responses [14]. There are a number of studies where bots with different capabilities – ranging from text-based chatbots, to more anthropomorphic virtual agents that can interact through voice – have been used to collect participant data [1, 12, 14, 21, 22, 27, 28, 31, 33, 35]. We summarise these studies to provide an overview of how bots can be used as a research tool to minimise the 'experimenter effect', highlighting benefits and current limitations of this approach.

### Reduce response biases

A recent study by Van Berkel et al. [33] investigated how crowdworkers evaluate the perceived fairness of predictors used in a recidivism algorithm, utilising 'SupportBot' to conduct the experiment without involving human researchers. 'SupportBot' provides a scalable mechanism to collect participants' private perceptions of fairness while simultaneously enabling public and structured discussions among participants (see Figure 1). The authors emphasise how the use of a bot for this study may have minimised possible experimenter effects when voting on the perceived fairness of sensitive predictors (*e.g.*, race of a defendant), which may have otherwise biased participant votes to be perceived positively by researchers [33].

Satisficing, *i.e.* the use of mental shortcuts to minimise effort required to complete a task, is a widely seen form of response bias among participants [14]. Although prior literature presumed that the absence of human experimenters may amplify such behaviour [17], a recent study by Kim et al. [14] note how a text-based chatbot with a casual conversational style (*e.g.*, friendly tone, emojis, punctuation) could facilitate sufficient conversational interactivity to reduce satisficing behaviour in online surveys. Kim et al. emphasise how conversing with the bot was perceived by respondents as a social interaction rather than a task. Moreover, the casual conversational style made participants attentive and engaged in their conversation, leading to high quality responses. The authors conclude that a text-based bot employing effective communication strategies can replicate certain abilities of a human interviewer.

**Figure 2: The virtual interviewer used by Lucas et al. [21] interacted with participants through speech.**

### Increase disclosure in sensitive topics

The perceived stigma associated with certain sensitive topics (*e.g.*, mental health) may induce limited disclosure and socially desirable responses from participants to avoid being negatively perceived by researchers [7]. In contrast, participants may feel more comfortable reporting negative self-admissions for sensitive topics to a virtual interviewer due to its anonymity and inability to judge them [28].

Recently, Lucas et al. [21] investigated how virtual interviewers (see Figure 2) can be used in clinical interviews to provide a 'safe' environment in which participants feel comfortable disclosing personal information. Participants were informed that they would interact with a virtual interviewer controlled by either a human operator or a computer. Participants were found to be more willing to disclose sensitive information when interacting with virtual interviewers perceived to be controlled by a computer, as compared to a human operator. Lucas et al. highlight that the perception of not being observed or judged by a human was sufficient to elicit more honest responses from participants.

Virtual interviewers have also been used in mental health research to increase reports of psychological symptoms by service members [22]. The study compared levels of disclosure among three conditions: an anonymous online symptom checklist, an interview with a human researcher, and an interview with a virtual interviewer. The highest number of symptoms were reported when participants were interviewed by virtual interviewers as a result of their ability to build relationships with participants while simultaneously allowing for anonymity.

### Facilitate cost effective scalability

As bots are not susceptible to human limitations such as fatigue, they can simultaneously interact with multiple research participants from across the world, with minimum resources, while also removing interviewer variability [27]. Thus, the cost effective scalability introduced by bots as a research tool could potentially mitigate the WEIRD problem in HCI research (*i.e.*, recruiting participants from Western, Educated, Industrialised, Rich, and Democratic countries), and increase the global applicability of study results by facilitating large-scale online experiments across countries and cultures [30]. Moreover, bots can be easily embedded in a diverse range of platforms such as social media networks [11, 14], smartphones [31], and virtual worlds [10], further enhancing its reach to potential participants.

Bots can also be used to perform tasks that a human experimenter may find challenging. For instance, Tallyn et al. [31] investigated the use of 'Ethnobot' to collect ethnographic data from multiple participants in the wild. Their findings highlight how 'Ethnobot' captured a variety of data (*e.g.*, participant activity, opinions, memories, doubts) simultaneously from groups of participants, in a short time period (26 hours of data in 2 days) which may be difficult for a single human ethnographer to achieve.

### Facilitate comprehension and clarification

The literature indicates that interacting with human experimenters allows participants to clarify doubts regarding the task at hand, while simultaneously allowing experimenters to request further clarifications from participants [12, 14]. Thus, the capacity of bots to facilitate a sufficient level of comprehension is crucial to be considered as an effective alternative research tool.

In a study by Bickmore et al. [1], a bot was seen to perform better than human experimenters in explaining research consent forms to potential participants, irrespective of a participant's literacy level. Participants were more likely to understand and sign the consent form when explained by a bot as the result of being comfortable to repeatedly ask questions from a computer. Moreover, a recent study evaluated the feasibility of using virtual interviewers to perform online cognitive assessments [35]. The authors proposed an 'Intelligent Interviewer' which can provide more accurate user assessments by offering participants the option to to ask for clarifications, and for the virtual interviewer to clarify conflicting responses through follow-up questions.

### Possible pitfalls

Using bots to minimise or completely remove human experimenters is not without limitations. As noted by Kim et al. [14], the ability to converse with participants in natural language (via text and/or speech) is why bots are superior to static online questionnaires. However, despite continuous advancements in NLP systems, bots are currently unable to gracefully handle complex and unstructured conversations [15]. As a result, bots are often restricted to structured conversations with limited flexibility which might make users feel less in control and frustrated [15, 31]. Moreover, despite reports on tone-aware chatbots that can respond empathetically to distressed users in customer service contexts [11], a fully automated bot is unable to replicate the richness of a human-to-human conversation. Especially in the case of mental health issues, researchers note how bots may not be able to fully understand the context of a respondent, and often provides off-topic or inappropriate responses which could undermine its ability to support patients in distress [16].

Finally, researchers highlight how virtual attributes (*e.g.*, skin colour) and/or speech style (*e.g.*, accent, pronunciation) of anthropomorphic bots could generate perceptions of its 'social identity' among respondents, as seen with human experimenters [4]. For instance, Krysan & Couper observe how photographs of virtual interviewers encouraged socially desirable responses, where White participants gave more racially conservative answers to virtual Black interviewers in comparison to White interviewers [18]. Similarly, gender [24] and personality [19] of a bot have also been observed to shape participant responses in diverse settings, thus threatening the effectiveness of anthropomorphic bots in mitigating experimenter effects.

**Advantages of using bots:**

- Reduces experimenter-induced response biases [33].
- Reduces satisficing and encourage engagement through conversational interactivity [14].
- Reduces fear of negative evaluation leading to more disclosure in sensitive topics [21].
- Provides a sense of anonymity and rapport [22].
- Facilitates cost efficient scalability [27].
- Provides diverse modes and platforms to interact with participants [31].
- Facilitates adequate comprehension and clarifications to participants [1].

**Current limitations of bots:**

- Restricted to structured conversations with limited flexibility [15, 31].
- Can't replicate human level support [16].
- Anthropomorphic bots may generate biases similar to humans [18, 24].

## CONCLUSION

This paper evaluated the benefits and limitations of using bots to mitigate experimenter-induced influences. We note that bots can facilitate conversational interactivity to adequately emulate human experimenters [14]. As a result, bots are a scalable and cost effective research tool [27] which can reduce biased responses [33] and enhance disclosure in sensitive contexts without compromising engagement and comprehension [1, 22, 35].

Our findings also indicate that a bot's design (*e.g.*, anthropomorphism, conversational capabilities) has significant implications on its effectiveness as a research tool. As users may regard bots as social actors [4], bots should be designed as to avoid inducing experimenter effects similar to humans. Alternatively, this notion could also be used for our benefit by customising bots to mirror participant demographics (*e.g.*, race, gender), to increase participant's willingness to disclose sensitive information and minimise stereotype threats [6, 23]. Another option would be to allow users to customise attributes of the bot (*e.g.*, appearance, speech style) based on their preference.

The use of bots for research also raises a number of ethical issues. While this is beyond the scope of this workshop paper, we point to recent work which has explored this aspect [10, 16]. In conclusion, while there is a substantial amount of research that supports the use of bots as a worthy alternative to human experimenters, more work is needed to explore its full potential.

## REFERENCES

[1] Timothy W Bickmore, Laura M Pfeifer, and Michael K Paasche-Orlow. 2009. Using computer agents to explain medical documents to patients with low health literacy. *Patient education and counseling* 75, 3 (2009), 315–320.

[2] Barry Brown, Stuart Reeves, and Scott Sherwood. 2011. Into the wild: challenges and opportunities for field trial methods. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1657–1666.

[3] Helen Chun, Maria I Tavarez, Grace E Dann, and Michael P Anastario. 2011. Interviewer gender and self-reported sexual behavior and mental health among male military personnel. *International journal of public health* 56, 2 (2011), 225–229.

[4] Frederick G Conrad, Michael F Schober, Matt Jans, Rachel A Orlowski, Daniel Nielsen, and Rachel Levenstein. 2015. Comprehension and engagement in survey interviews with virtual agents. *Frontiers in Psychology* 6 (2015), 1578–1798.

[5] Ronald Cosper. 1972. Interviewer effect in a survey of drinking practices. *The Sociological Quarterly* 13, 2 (1972), 228–236.

[6] Darren W Davis and Brian D Silver. 2003. Stereotype threat and race of interviewer effects in a survey on political knowledge. *American Journal of Political Science* 47, 1 (2003), 33–45.

[7] Rachel E Davis, Mick P Couper, Nancy K Janz, Cleopatra H Caldwell, and Ken Resnicow. 2010. Interviewer effects in public health surveys. *Health education research* 25, 1 (2010), 14–26.

[8] Nicola Dell, Vidya Vaidyanathan, Indrani Medhi, Edward Cutrell, and William Thies. 2012. "Yours is better!" participant response bias in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1321–1330.

[9] Pamela Grimm. 2010. Social desirability bias. *Wiley international encyclopedia of marketing* (2010).

[10] Béatrice S Hasler, Peleg Tuchman, and Doron Friedman. 2013. Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds. *Computers in Human Behavior* 29, 4 (2013), 1608–1616.

[11] Tianran Hu, Anbang Xu, Zhe Liu, Quanzeng You, Yufan Guo, Vibha Sinha, Jiebo Luo, and Rama Akkiraju. 2018. Touch your heart: a tone-aware chatbot for customer care on social media. In *Proceedings of the 2018 CHI Conference on Human*

*Factors in Computing Systems*. 1–12.

[12] Timothy P Johnson, Michael Fendrich, Chitra Shaligram, Anthony Garcy, and Samuel Gillespie. 2000. An evaluation of the effects of interviewer characteristics in an RDD telephone survey of drug use. *Journal of Drug Issues* 30, 1 (2000), 77–101.

[13] Sina Kianersi, Maya Luetke, Reginal Jules, and Molly Rosenberg. 2019. The association between interviewer gender and responses to sensitive survey questions in a sample of Haitian women. *International Journal of Social Research Methodology* (2019), 1–11.

[14] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[15] Lorenz Cuno Klopfenstein, Saverio Delpriori, Silvia Malatini, and Alessandro Bogliolo. 2017. The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In *Proceedings of the 2017 conference on designing interactive systems*. 555–565.

[16] Kira Kretzschmar, Holly Tyroll, Gabriela Pavarini, Arianna Manzini, Ilina Singh, and NeurOx Young People's Advisory Group. 2019. Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical informatics insights* 11 (2019), 1–9.

[17] Jon Krosnick. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology* 5, 3 (1991), 213–236.

[18] Maria Krysan and Mick Couper. 2006. Race of Interviewer Effects: What Happens on the Web? *International Journal of Internet Science* 1, 1 (2006), 17–28.

[19] Jingyi Li, Michelle X Zhou, Huahai Yang, and Gloria Mark. 2017. Confiding in and listening to virtual agents: The effect of personality. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. 275–286.

[20] Mingnan Liu and Yichen Wang. 2016. Race-of-Interviewer Effect in the Computer-Assisted Self-Interview Module in a Face-to-Face Survey. *International Journal of Public Opinion Research* 28, 2 (2016), 292–305.

[21] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100.

[22] Gale M Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. 2017. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI* 4 (2017), 51.

[23] David M Marx and Phillip Atiba Goff. 2005. Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology* 44, 4 (2005), 645–657.

[24] Marian McDonnell and David Baxter. 2019. Chatbots and gender stereotyping. *Interacting with Computers* 31, 2 (2019), 116–121.

[25] Martin T Orne. 1962. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist* 17, 11 (1962), 776–783.

[26] G. Payne and J. Payne. 2004. The Hawthorne Effect. *Key Concepts in Social Research* (2004), 108–111.

[27] Matthew Pickard, Ryan M Schuetzler, Joseph Valacich, and David A Wood. 2019. Innovative accounting interviewing: A comparison of real and virtual accounting interviewers. *The Accounting Review, Forthcoming*. (2019), 69.

[28] Matthew D Pickard, Catherine A Roster, and Yixing Chen. 2016. Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions? *Computers in Human Behavior* 65 (2016), 23–30.

[29] M. Schober and F. Conrad. 1997. Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly* (1997), 576–602.

[30] Christian Sturm, Alice Oh, Sebastian Linxen, Susan Dray, and Katharina Reinecke. 2015. How WEIRD is HCI? Extending HCI Principles to other Countries and Cultures. In *CHI'15 Extended Abstracts on Human Factors in Computing Systems*.

2425–2428.

[31] E. Tallyn, H. Fried, R. Gianni, and C. Speed. 2018. The Ethnobot: Gathering Ethnographies in the Age of IoT. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[32] Aditya Vaidyam, Hannah Wisniewski, John Halamka, Matcheri Kashavan, and John Torous. 2019. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian Journal of Psychiatry* 64, 7 (2019), 456–464.

[33] Niels van Berkel, Jorge Goncalves, Danula Hettiachchi, Senuri Wijenayake, Ryan M Kelly, and Vassilis Kostakos. 2019. Crowdsourcing Perceptions of Fair Predictors for Machine Learning: A Recidivism Case Study. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–21.

[34] Lisa Slattery Walker. 2014. Developing Your Experiment. In *Laboratory Experiments in the Social Sciences (2nd Edition)*. Academic Press, 127 – 144.

[35] Kathleen Weaver, Anita Komlodi, and Brian Duffy. 2013. Using an intelligent interviewer to perform cognitive assessments. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 259–264.

[36] Senuri Wijenayake, Niels van Berkel, Vassilis Kostakos, and Jorge Goncalves. 2019. Measuring the Effects of Gender on Online Social Conformity. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 145 (Nov. 2019), 24 pages.

[37] Senuri Wijenayake, Niels van Berkel, Vassilis Kostakos, and Jorge Goncalves. 2020. Impact of contextual and personal determinants on online social conformity. *Computers in Human Behavior* 108 (2020), 106302.

[38] Sandra R Wilson, Nancy L Brown, Carolina Mejia, and Philip W Lavori. 2002. Effects of interviewer characteristics on reported sexual behavior of California Latino couples. *Hispanic Journal of Behavioral Sciences* 24, 1 (2002), 38–62.

[39] Rainer Winkler and Matthias Soellner. 2018. Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. In *Academy of Management Proceedings*, Vol. 2018. 15903.