Dynamic Voice Clones Elicit Consumer Trust

Scott Schanke
University of Wisconsin-Milwaukee
Lubar School of Business
schanke@uwm.edu

Gordon Burtch
Boston University
Questrom School of Business
gburtch@bu.edu

Gautam Ray
University of Minnesota
Carlson School of Management
gautamr@umn.edu

Abstract

Platforms today are experimenting with many novel personalization technologies. We explore one such technology here, voice-based conversational agents, with a focus on consumer trust. We consider the joint role of two key design / implementation choices, namely i) disclosing an agent's autonomous nature to the user, and ii) aesthetic personalization, in the form of user voice cloning. We report on a set of controlled experiments based on the investment game, evaluating how these design choices affect subjects' willingness to participate in the game against an autonomous, AI-enabled partner. We find no evidence that disclosure affects trust. However, we find that the greatest level of trust is elicited when a voice-based agent employs a clone of the subject's voice. Mechanism explorations based on audio analytics indicate that voice-cloning induces trust by eliciting a perception of homophily.

1. Introduction

Voice-based autonomous agents, such as Google Assistant, Siri, and Alexa, are playing an increasingly prominent role in the digital economy. Some industry estimates suggest that 38.5% of Americans interact with voice-based assistants, and Gartner has projected that the current market for Voiced-based Personal Assistants will soon reach $3.5 Billion, comprised of both consumer and employee support applications. As this technology begins to permeate work and day-to-day life, several digital providers and platform operators are beginning to explore the design of voice-based agents, to create engaging interfaces, and optimize the user experience.

Voice-based AI is not only top of mind for platform operators; governments and regulators have also turned their attention to these new technologies. As a primary example, the California State legislature recently passed the Bolstering Online Transparency (B.O.T.) Act (aka the "blade-runner bill") which requires that firms disclose the autonomous nature of conversational customer service agents at the outset of any customer interaction. Fines for undisclosed agents can cost an organization up to $2,500 per interaction in the state of California. Additionally, it may only be a matter of time before the U.S. federal government requires the same amount of transparency as the Federal version, the Bot Disclosure and Accountability Act of 2018, recently landed on the Senate floor.[1]

Although the regulatory transparency requirements help to protect consumers, disclosing an agent's autonomous nature may have severe consequences for consumer perceptions and interactions. For instance, in specific contexts like movie and book recommendations [1], or medical diagnosis [2], individuals have been shown to distrust algorithms. However, in other contexts, such as forecasting, dating recommendations [3], or investment advising [4], individuals have often been found to prefer algorithms. Accordingly, depending on the organizational context, there may be benefits or detriments to disclosing the true nature of an autonomous agent.

While literature on algorithmic aversion and appreciation has considered a variety of contexts, relatively little work has considered e-commerce applications, despite that these are likely to be one of the most common use-case for conversational autonomous agents in the coming years. Exceptions include the work by [5], who conducted a study of voice-based chatbots working on behalf of a financial institution, soliciting loan renewals from existing customers. Those authors found that disclosure of a chatbot's autonomous nature could lead to significant reductions (~79.7%) in conversion, depending on whether the disclosure was made up-front or after some delay [5]. In a digital retail context, [6] reported that disclosure of a text-based chatbot's autonomous nature drove significant increases in perceived likeability. Given this contrast in prior results, it remains relatively unclear by what mechanism disclosure influences consumer perceptions, nor how

---

[1] https://www.congress.gov/bill/115th-congress/senate-bill/3127

disclosure may affect consumer trust in a more general sense. We therefore revisit this question, considering the implications of disclosure for trust, a fundamental aspect of economic exchange that is crucial to many consumer contexts [7]. Disclosure is likely to influence trust because the decision to disclose may foster transparency on the part of a user. Further, failure to explicitly disclose may have negative consequences if a user independently perceives the autonomous nature of the other party.

With the changing legal climate moving in the direction of forced disclosure, organizations may require ways to mitigate the negative appraisals of a fully disclosed agent. One potentially fruitful avenue is personalization. More specifically, a long stream of work speaks to the relationship between homophily and trust, including as it relates to voice-based similarity. While voices have several features[8], research in psychology has found that individuals are more attracted to others when they share similar voice characteristics [9]. Thus computer-generated voices are preferred when their features imply personality traits that are similar to the user [10]. These prior findings are notable because recent technological innovations enable the possibility of voice cloning, in near real-time, on the basis of relatively small voice samples [11]. With this new technology, there now exists the possibility that organizations can dynamically personalize a voice-based autonomous agent to a particular user .

Furthermore, it is essential to consider how disclosure may interact with dynamic personalization. More simply, in the presence of disclosure, we might anticipate that personalization will be perceived positively. For instance, agents that use disclosure may be approached with less scrutiny or cynicism by the user; conversely, absent disclosure, personalization attempts may trigger an adverse reaction if perceived as duplicitous or manipulative. Thus, we explore the following formal research questions in this work: ***To what extent does voice cloning induce trust in an autonomous (spoken) agent? Does disclosure of a spoken agent's autonomous nature causally impact user trust? How does disclosure affect the response to voice-based personalization?***

To address these questions, we build on a version of the behavioral economics "trust game" implemented by [12, 13], whose studies examined the role of communication on trust. In our adaptation, two players, a human subject and an AI agent, play a one-shot game with each other. After pairing with the agent, the human subject faces a choice of whether to trust the other party. Suppose the human player decides to trust the agent. In that case, the player faces the risk

that the automated player is designed to act selfishly, maximizing its payoff. Under this setup, two aspects of the interaction are experimentally manipulated. First, we randomly assign some subjects to AI disclosure; that is, for a random set of subjects, at the outset of the pairing, the AI agent's autonomous nature is explicitly disclosed. Second, we randomize human subjects to receive alternative forms of communication from the AI agent (or no communication, in the Control condition). Thus, before the human subject makes a decision about whether to play, i.e., opting-in, in some conditions the automated agent issues a message to the subject. This message is communicated either textually or verbally. In verbal conditions, the voice conveying the message is either a dynamic clone of the subject's voice or it is a default male voice. These two manipulations allow us to understand how the agent's voice influences the subject's behavior, and how that influence varies with explicit AI disclosure.

This game set up is relevant for consumer facing AI for two reasons. First, in implementations of AI in human facing job roles, agents act as a subordinate [14], working on an organizations behalf, assisting the consumer. For instance, a customer may call-in to a company's customer service-line and require the automated agent to perform some task (e.g. change a flight booking, update dinner plans, or perform a bank transfer). In a scenario such as this, there is a degree of uncertainty as to whether the agent will indeed perform as expected. Our game is an appropriate representation for this real world context, as [12, 13] designed to study scenarios with hidden-information, where true actions to be taken by an agent are unknown. Secondly, the game scenario used in this setting focuses on an economic based version of trust. While there are several versions of trust used in designing technology artifacts like [15], our version of trust in this game provides a more realistic view of how individuals will make real-time decisions. Much aligned with [16], the use of this behavioral game, will most accurately mirror human decision making as opposed to using imagined vignettes.

As we will demonstrate, subjects are most apt to participate, and thus trust, when they engage with a voice-based agent that is explicitly disclosed, and which bears a voice signature that is a clone of the subject's own voice. Interestingly, mere disclosure is not sufficient to significantly impact subject trust. We speculate here that the relative lack of trust exhibited in other conditions, even with disclosure, may be driven by subjects' discerning or suspecting that their counterpart is indeed automated, regardless of explicit disclosure. When the agent is not disclosed, the subject may interpret this as an attempt at deception that harms the

appraisals of trustworthiness [17].

Our findings are important as they contribute to both research and practice. From a theoretical standpoint, this work contributes to the burgeoning work assessing human acceptance of algorithms [5, 18, 19, 20] by critically examining how and why individuals trust algorithms in an economic scenario, and how the aesthetic aspects of these systems can influence behavior and perception. Second, we contribute to the literature on personalization in information systems [21] and marketing [22, 23]. We seek to evaluate whether and to what extent a new personalization technique, dynamic voice cloning, can influence consumer perceptions and engender trust. Finally, our research aims to inform practice, as we investigate how AI disclosure, a regulatory requirement on the horizon, influences human interaction with audio-based conversational agents. Further, we describe an A/B testing framework for evaluating the design and implementation of voice-based AI systems.

2. Literature Review

2.1. Algorithms in Human Facing Job Roles

The use of algorithms to support decision-making is a lengthy line of inquiry [24, 25, 26]. As algorithms have become the bedrock of our society, it is no surprise that human interaction with algorithms is an emerging topic in Marketing [4, 5], Psychology [27, 3], Computer Science [28], and Economics [29].

One specific stream of research, of relevance here, focuses on human appreciation and aversion to algorithmic judgment. While this stream of work is developing, much of the findings categorize human perceptions of algorithms as highly contextual. In some cases, individuals appreciate algorithms, specifically when dealing with concrete, objective outcomes like numeric forecasts [3], providing geographic directions, or predicting the weather [4]. In other cases individuals tend to prefer human judgment, particularly for more subjective tasks, like music recommendation [28], HR decisions [4, 20], and medical diagnoses [2]. As human-facing job roles comprise a mix of objective and subjective tasks, it is possible that individuals may have positive or negative reactions to algorithms in these contexts.

Aligned with the notion that humans react negatively toward AI in these roles, [5] find that customers dislike interacting with AI agents in a sales role. More specifically, they executed a field experiment with a financial loan company in China. A voice-based AI agent placed a sales call and asked customers whether they would like to sign up for a loan renewal [5]. In outbound calls, the authors randomly manipulate whether to disclose that the agent is autonomous [5]. This simple act of disclosure reduced loan renewal by 79.7% [5]. These authors' findings reinforce that there could be drastic real-world consequences to organizations if disclosure is a regulatory requirement.

In contrast, other authors have found that humans can be receptive to AI in these roles. In a lab experiment, [6] manipulated disclosure of a highly anthropomorphized text-based AI Agent in a customer service setting. There we find that disclosure of an AI agent's autonomous nature significantly improves its likeability. Additionally, [18], evaluate how workers in a sales setting react to human versus AI sales coaches. In a series of field experiments, sales agents work with potential customers to sell loans and are randomly assigned to receive a human versus AI sales coach [18]. Notably, those authors observed sales improvement for the average salesperson receiving the AI coach instead of the human [18].

As the emerging literature in human interactions with algorithms in human-facing job roles is fractured and likely dependent on a myriad of contextual factors, a more general consideration is needed to enable our understanding of how and when human's may trust algorithms in an e-commerce setting. Experimental economics is replete with designs that have been developed to assess fundamental human behavior in response to basic stimuli, to identify and examine human decision-making biases and strategic behavior (e.g. trust, cooperation, negotiation). We thus draw on this literature, in an effort to further our understanding of how humans trust and react to algorithms [30, 31, 32]. While the prior literature in behavioral economics has not focused on studying human-machine interactions, the field has inadvertently studied this phenomenon on many occasions, through the use of computer players in experimental games [31, 33, 34]. Notably, this is also true of some work in Human Computer Interaction (HCI) (e.g., [35, 36, 37]) and Organizational Behavior [19].

When individuals interact with computer players in economic experiments, a common finding has been that subjects tend to behave more rationally and act more for their personal gain [19, 31, 34]. More specifically, individuals negotiate less emotionally with computer players [19, 38, 34], they more readily cooperate [39], they bid more conservatively in auctions [40], and they are more likely to exploit learned behavior in repeated play [41]. The reasoning for the altered behavior relative to experiments involving human partners is that there is a reduced role for social considerations when subjects

interact with computer players [31].

While most of these interactions exist through a simple computer interface, in the rare instance where subjects have faced a computer player that possesses human-like qualities, subjects' behavior has been notably less predictable. For instance, in a prisoner's dilemma game, human subjects are more likely to reject cooperation with a computer player imbued with a human-like voice compared, compared to a text-based computer or human [35]. Additionally, individuals tend to reason through offers similarly when given by a human-like robot or by a true human [36]. These findings are particularly applicable to settings wherein AI-enabled agents operate in human-facing job roles, as their implementation likely utilizes some form of human-like qualities, like voice interaction [5] or communicative dialog and social cues. As humans tend to anthropomorphize AI agents [42], one way to design persuasive interfaces is to look to the social variables identified in human-to-human interactions [37].

2.2. Similarity Attraction

Similarity attraction is a phenomenon that has been studied by social psychologists for decades [43]. This type of attraction occurs when individuals appraise others as being similar to themselves in some personal dimension like facial attributes [44], demographics [45, 46], and attitudes [47, 48]. Although initially observed in face-to-face interpersonal contexts, similarity-based attraction transfers to several online contexts: social media [49, 50], computer-mediated communication [51], and voice-based exchanges [9] to name a few.

HCI designers have also noticed human affinity towards AI agents that are similar in some respects to the focal user. More specifically, individuals trust AI agents that possess similar paralinguistic vocal cues that align to their personality [10] and display accents that are congruent with their own [9]. In a recent experiment investigating human AI teamwork, [52] found that people prefer to work with AI agents whose avatar image displayed similar racial and gender characteristics and ultimately trusted these agents more than those that were dissimilar. As they relate to AI Agent design, these findings show that individuals tend to trust systems that display similar aesthetics to themselves.

Although there are many reasons for similarity attraction, a dominant explanation for the phenomenon is the reinforcement-affect model [53]. This model postulates that people enjoy experiencing positive stimuli [53] and actively seek them. Since interactions with similar individuals tend to go more smoothly, as opposed to dissimilar stimuli, individuals are drawn toward stimuli congruent with these attributions [54]. For example, when individuals hear a similar voice, they associate the sound with subsequent positive conversations, as there is already a basis for common ground. That said, it is somewhat unclear how disclosure of an AI agent will interact with similarity-based aesthetics. On the one hand, disclosure alone may be enough to push individuals away from the AI agent as it highlights dissimilarity. However, on the other, it may mitigate the initial negative dissimilarity appraisals.

As there are reasons to believe that AI agent disclosure will influence trust, and similar voice aesthetics may mitigate negative aspects invoked by disclosure, we look to a behavioral economics experiment to help us evaluate these features.

3. Methods

3.1. Experiment Design

We explore the influence of disclosure and voice-similarity on trust by utilizing an experiment design first introduced by [12], which was specifically developed to understand how communication, generally, influences trust. In the original game design, articulated in Figure 1, subjects are first "matched" to a human playing partner and randomly assigned to play the role of agent A or B. In the original treatment condition, subjects playing the role of B would have the opportunity to pass a written message to party A, stating whatever they like, e.g., superfluous comments, promises, etc. Subsequently, party A would decide whether they wished to play the game with B (choosing "in" or "out"). If party A decided to play, B could choose to walk away with some money, leaving A with nothing, or B could choose to roll a die. Choosing to roll, the final payoff to both parties would depend on the resulting number that was rolled, with a 5 in 6 chance of a positive payoff for both players. This study design enabled [12] to examine how communication would influence the decisions of both A (whether to trust B), and B (how to behave upon receiving that trust).

We modified this game, in three key ways. First, we re-implemented the game in a digital setting, incorporating voice-based messaging in addition to the original text-based communication that was considered in the original study. Second, we assign party B to always be played by an autonomous agent. We thus have two experimental conditions that generally mirror those of the original study design: a control condition, where no communication would take place, and a written (text-based) message condition, wherein
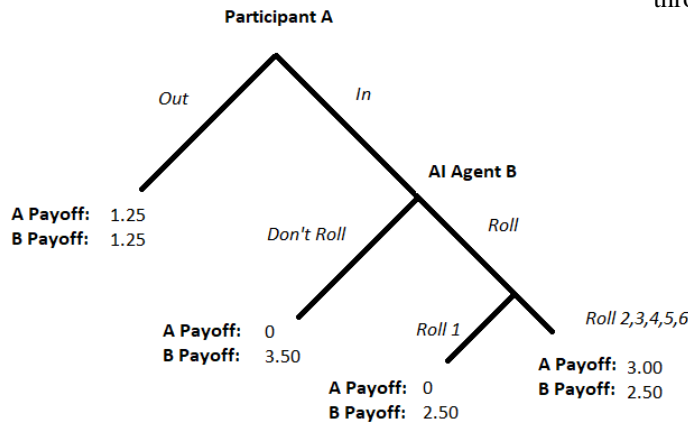
Figure 1. Payoffs Adapted from Charness & Dufwenberg 2006

through their computer, via a web interface.

Figure 2. Study Enrollment Process

party A would receive a written text-based message from party B. In the latter case, for the sake of realism, we simply reused messages at random that were drawn from those exchanged by subjects in the original studies as documented by [12, 13]. These messages were mostly promises of rolling the dice if the subject chose in. Finally, we also fixed the agent's choice where if the subject opts in, the agent always chooses to roll. There is no breaking of promises in our version.

We then supplement the above conditions with others of our own devising. We added conditions wherein the communication would take place via a voice-based message, i.e., a recording. Further, we manipulated the voice used in generating that recording, employing a 'default' computer voice in one condition and a dynamic voice-clone in another condition (we generate this clone based on a consent statement recorded via telephone prior to the game beginning - we provide more details on our procedure below). Finally, we incorporate a set of additional conditions that mirror the voice-based conditions above, as well as the control condition, with the simple added element that we disclose the autonomous nature of Party B to the subject. This leaves us with 7 experimental conditions: Control, Text, Control + Disclosure, Default Voice, Default Voice + Disclosure, Clone Voice, and Clone Voice + Disclosure.

A detailed outline of how subjects play our game can be found in Figure 2, which depicts the three-step process. First, a subject would call into a 1-800 number and reads a consent statement over the phone. Next, a server records the consent statement, randomizes the subject into one of the experimental conditions, and returns a unique completion code. Third, the subject then enters the completion code and plays the trust game
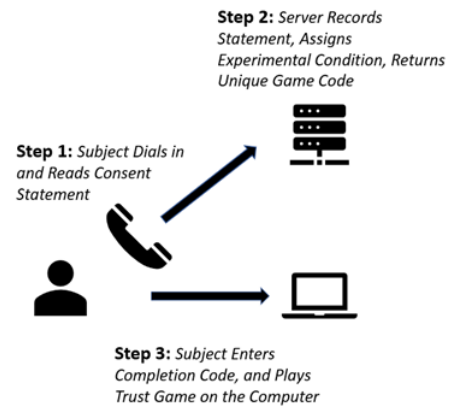
3.2. Voice Cloning

To implement voice cloning, we utilized an open-source implementation of [11]. This voice cloning method uses transfer learning, a Deep Learning technique, which allows for learning in one task to transfer over to another [55]. The voice cloning model is comprised of three parts: a Speaker Encoder, Synthesizer, and Vocoder. The Speaker Encoder takes a short audio file and generates a speaker embedding. The embedding is used with the Synthesizer, which maps the to-be-generated spoken words to a mel-spectrogram. This spectrogram is fed into the Vocoder, which takes the spectrogram image and generates an audio file. An illustration of [11], 2018's three-part model can be found in Figure 3 .
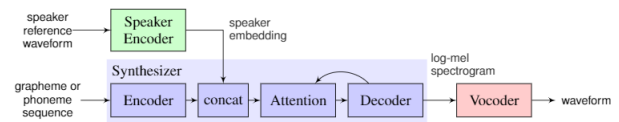

Figure 3. Jia et al 2018 Voice Cloning Model

The voice cloning software was installed on an Amazon Web Services GPU instance. When a subject, randomized into a voice clone condition, reads their consent statement, this pipeline is triggered. The resulting audio is loaded onto an S3 bucket, which would then be presented as a message communicated by Player B during game play. Party A would receive this message just prior to making their decision about whether to opt-in, to play with Party B. Similarly,

disclosure would always take place prior the opt-in decision.

4. Empirical Approach & Data

Our outcome variable of interest is whether the subject chooses to be "in," trusting player B sufficiently to play the game. Opt-in is thus reflected by a binary indicator variable. Additionally, we utilize a series of treatment dummies to capture each experimental condition. For the analysis, we examine group-specific opt-in rates, and we execute logistic regression and Fishe Exact tests to identify how these treatments associate with opt-in.

4.1. Subject Recruitment

Before conducting the experiment, a power analysis was conducted, assuming effects consistent with those observed by [12], in terms of sign and magnitude. We recruited 1,118 subjects through Prolific. We purposefully over-allocated subjects to the control conditions (i.e., Control and Control + Disclosure) to reduce load on our voice-generating servers, as generating too many audio messages in a short period would cause messages to be generated too slowly for the experiment to progress correctly. Additionally, we limited the subject pool to individuals from the United States and Canada, as the original AI models used in producing the recorded messages were generated based on audio from North American English speakers. Limiting the subject pool in this way prevented additional accent based similarity or location attributions [9].

4.2. Descriptive Statistics

Overall, shown in Table 1, participants trusted player B on average 75.4% of the time. They also spent 322 seconds (about 5 minutes) playing the game, on average.

Table 1. Descriptive Statistics

Variable	Mean	St. Dev.	Min	Max
In-Rate	0.754	0.431	0.00	1.00
Total Time (sec.)	322	194	48	263
Disclosure	0.410	0.492	0.00	1.00
Message	.493	0.50	0.00	1.00
Age	33.4	11.8	18	78

We report the number of subjects per condition in Table 2. As noted above, more subjects were allocated into the control conditions to reduce load on our servers handling voice-synthesis.

Table 2. Subjects per Condition

Condition	Obs.
Control (No Message) – CD 2006	331
Text – CD 2006	109
Control + Disclosure	236
Default Voice	108
Default Voice + Disclosure	107
Cloned Voice	112
Cloned Voice + Disclosure	115

5. Results

5.1. Voice Cloning and Disclosure's Independent Influence on Opt-In

Overall results of our experiment by message type is shown in **Table 3**. Model free evidence would suggest that opt-in rate for voice-cloning is higher than both the Default Voice and No Message conditions. As compared to the text-based *Charness and Dufwenberg 2006* conditions, Cloning appears to perform at least as good as the undisclosed text-based condition. To determine the influence of voice cloning on opt-in rate, we estimate a logit model with the binary outcome variable *Opt In*, and the condition Clone as the reference condition. As shown in **Table 3**, we find that Cloning an end users voice does indeed encourage higher opt-in from the subject as compared to *No Message* and the *Default* voice message.

$$OptIn_i = \alpha + \beta_1 \cdot 1 \, MessageType_i + \epsilon_i \quad (1)$$

Table 3. Message Type (Logit)

Variable	DV = Opt-In
$NoMessage$	-1.0787^{***} (0.036)
$Default$	-0.5150^{**} (0.034)
$Text$	-0.2888 (0.055)
$Intercept$	0.108^{**} (0.055)
$Observations$	1,118

Note: $^{**} \, p < 0.05, \, ^{*} \, p < 0.10.$

Somewhat surprising, when comparing Opt-In for *Undisclosed* and *Disclosed* partners, rates appear identical at (75%). In conducting pairwise χ^2 tests we find no statistically significant difference between the two conditions (p>.1). Thus we can conclude that, at least in our context, individuals do not have a bias towards interacting with humans over AI agents.

5.2. Voice Cloning and Disclosure

We next report a series of pairwise Fisher-Exact tests demonstrating the relative performance of all seven conditions. Figure 4 depicts the opt-in rate for each condition. Here we can see that the Clone + Disclosure condition achieves the highest opt-in rate (89.5%), and thus appears to garner the greatest level of trust. To
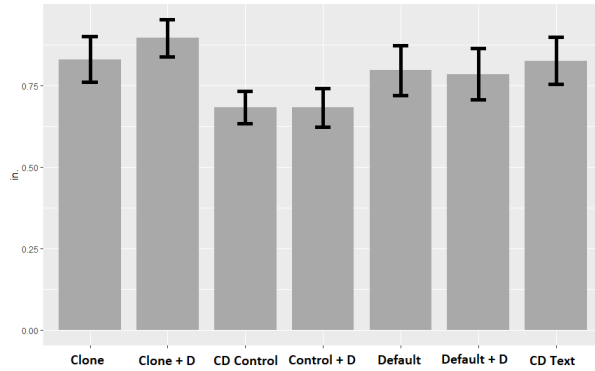

Figure 4. In Rate by Condition - (95% Confidence Intervals)

assess this result, we conduct pairwise comparisons of each of these conditions through use of the Fisher Exact test. Interestingly, the Clone + Disclosure condition has a significantly higher Opt-In rate as compared to Disclosed or Undisclosed No Message and Default Message conditions, this is significant at the (p<.01) and (p<.05) respectively [2]. Additionally, there is no significant difference between and among both Clone conditions and the undisclosed text condition. This lends support for use of voice cloning as it compares to a default voice, however the clone performs no better than the text based condition.

5.3. Randomization Checks

In our experimental setup, we randomized individuals when they entered the experiment. To ensure that the randomization process occurred successfully, we assess balance across conditions in terms of several observable characteristics of subjects including age, latitude, longitude and gender and educational attainment. As shown in Table 4 and Table 5, we observe no significant imbalance on any of these variables. Although these linear regressions effectively report pairwise comparisons with the control condition (taken as reference), the results remain

[2]These p values do not utilize an adjustment for false discovery, when this is incorporated the p-values are slightly higher with (p<.05) and (p<.1)

similarly null when we employ alternative conditions as reference.

Table 4. Randomization Balance Checks (OLS Regression)

	Female	Lat	Lon	Age
Control-D	0.03	−0.63	1.27	0.34
	(0.04)	(0.45)	(1.40)	(0.99)
Default	−0.07	0.67	−0.73	−1.47
	(0.05)	(0.58)	(1.82)	(1.28)
Default-D	0.04	0.12	0.40	0.05
	(0.05)	(0.59)	(1.85)	(1.30)
Text	0.02	0.19	1.70	0.65
	(0.05)	(0.59)	(1.83)	(1.29)
Clone	0.00	0.27	1.49	−1.11
	(0.05)	(0.58)	(1.81)	(1.27)
Clone-D	0.04	0.46	1.85	−0.61
	(0.05)	(0.57)	(1.77)	(1.25)
(Intercept)	0.44***	37.65***	−92.64***	33.75***
	(0.03)	(0.29)	(0.91)	(0.64)
R^2	0.00	0.01	0.00	0.00
Adj. R^2	−0.00	0.00	−0.00	−0.00

$^{***}p < 0.01;\ ^{**}p < 0.05;\ ^{*}p < 0.1$

Table 5. Randomization Check (MLOGIT; DV=Message Type)

Variable	No Message	Default	Text
Bachelor's	-0.059	-0.354	-0.045
	(0.059)	(0.271)	(0.315)
High school	0.435	-0.137	-0.900
	(0.340)	(0.415)	(0.617)
Less than high school	-0.064	-0.481	0.141
	-(0.734)	(0.940)	(0.951)
Some college	0.306	0.091	-0.465
	(0.243)	(0.280)	(0.359)
Constant	-0.039 ***	1.27e-08	-0.019 *
	(0.089) ***	(0.088)	(0.087)
Observations		1,118	
Pseudo R^2		0.0076	
Chi^2		20.636,(p >.05)	

$^{***}p < 0.01;\ ^{**}p < 0.05;\ ^{*}p < 0.1$

6. Mechanism Exploration

Next, to better understand the mechanisms that underlie our results in the main experiment, we ran a follow-on experiment, employing a similar design. However, instead of testing disclosure, we seek to understand which features of an AI agent's voice may be driving the increased trust by the subject. In this experiment we have three conditions: 1) Cloning with Disclosure, 2) Default Male Voice with Disclosure 3) Random Voice with Disclosure. The Random Voice with disclosure is a random draw of speaker embeddings from the Librispeech dataset. [3]

To analyze how individual features influence trust, we execute a causal mediation analysis [56]. For the

[3]https://www.openslr.org/12

Table 6. Subjects per Condition

Condition	Obs.
Clone + Disclosure	221
Default Voice + Disclosure	219
Random Voice + Disclosure	226

analysis of vocal features, we utilize PRAAT [4], an open source computer tool commonly used in phonology. We generate features associated with both the phone-read consent statement and the AI generated audio messages. We construct a normalized score of four different features that are typically used to characterize voice signatures in spoken language, namely fundamental frequency (f0), harmonic to noise ratio (hnr), vocal jitter, and vocal shimmer. Based on the normalized difference among these four features, between the generated audio and the human subject's recorded audio, we create an index / score reflecting the voice-similarity between each subject and their voice-based, AI-enabled partner. Figure 5 illustrates the mediated relationship of vocal differences on trust.
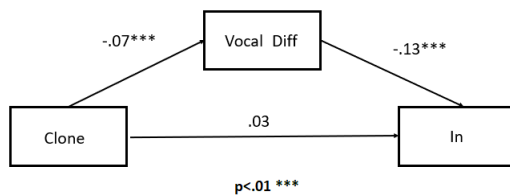

Figure 5. Vocal Differences Indirect Influence on Trust

As shown, there is a negative relationship between clone and vocal differences. This makes sense as cloning is meant to reduce the distance between the speaker and the generated voice. Secondly, we see that the vocal difference is also negatively associated with trust. This would suggest that the closer the match to the focal speakers voice, the more likely the subject will choose in. This strengthens the argument for voice clones, or similar sounding voices being used to help influence trust.

7. Discussion

Our work provides a novel first look at a potentially impactful design aesthetic of spoken AI agents, dynamic voice cloning. Our behavioral economics experiment

[4]https://www.fon.hum.uva.nl/praat/

shows that incorporating voice cloning may be a viable option to elicit trust from individuals interacting with AI agents in collaborative and economic interactions.

Secondly, this work evaluates how AI Agent disclosure influences trust in these artifacts. While previous fieldwork suggests that individuals react adversely when agents disclose that they are automated [5], from our study, we find that disclosure by itself does not necessarily negatively impact trust. That said, some of the discrepancies between [5], could be that the "realism" of voices utilized in our experiment is lower than theirs. We used recordings from lower quality telephone calls, thus making the generated output somewhat rough and less lifelike. In turn, users may readily identify that the agent is not a human and breaking any initial trust perceptions a subject may have. Regardless of the voice, mending this broken trust may prove difficult [17]. It may also be a cautionary tale for organizations that look to pursue a non-disclosure strategy, as it may destroy consumer trust at the onset.

Perhaps a more realistic comparison to a true human actor, in our experiment, would be the text-based CD condition, as text messaging has few social cues for subjects to determine if the agent is indeed human [57, 58]. That said, our dynamic voice clone with disclosure condition performs at least as well as the CD condition. This result furthers the notion that a potentially more viable strategy for organizations to pursue is voice-based personalization with disclosure, as there is no benefit to withholding this information.

Thirdly, we evaluated how disclosure interacts with dynamic voice cloning. From our results, we see that disclosure, when paired with voice cloning, achieves the highest in rate to a statistically significant degree over the default voice (with or without disclosure) and no message conditions. These findings are interesting as it shows a potential path forward for organizations if faced with increased regulatory disclosure requirements. Additionally, further data requirements and regulations, outside of AI disclosure, like GDPR, could be on the horizon , limiting what archival data organizations can use to personalize their customers' experience [59]. Dynamic voice cloning can be done on the fly with a small clip of a customer's voice without needing to persist any user data.

While our research evaluates a potential new way for organizations to personalize their audio-based experiences, our work has several limitations. Firstly, our experimental context occurs in a controlled online lab experiment; it is unclear how much trust is involved in outcome variables of interest for marketers and organizations (e.g., sales conversions, lead generation). Although we feel that the lab-based setting gives our

research the ability to tease out the potential influence of disclosure on trust, further work should be done in the field evaluating the efficacy of dynamic voice cloning in the field.

Another limitation of our work is its focus on North American English-speakers. It is not clear if our results necessarily generalize across cultures. For example [5] which takes place in a field setting in China, found starkly different results as they relate to disclosure. Further work on AI Agent disclosure could seek to evaluate if there are perhaps country-based differences in the reaction to this feature.

8. Conclusion

In conclusion, this work seeks to help AI practitioners evaluate two crucial features, AI disclosure, and voice-based aesthetic. While one feature, AI Disclosure, may be a regulatory requirement for all AI Agents in the future, we find that dynamic voice cloning may help to elicit higher levels of trust in disclosed voice-based systems. As voice-based AI is becoming an ever more critical consumer channel, dynamic voice cloning could be a potentially fruitful step forward in designing transparent voice-based consumer experiences.

References

[1] K. Swearingen and R. Sinha, "Beyond Algorithms : An HCI Perspective on Recommender Systems," *ACM SIGIR 2001 Workshop on Recommender Systems (2001)*, pp. 1–11, 2001.

[2] M. Promberger and J. Baron, "Do patients trust computers?," *Journal of Behavioral Decision Making*, vol. 19, no. 5, pp. 455–468, 2006.

[3] J. M. Logg, J. A. Minson, and D. A. Moore, "Algorithm appreciation: People prefer algorithmic to human judgment," *Organizational Behavior and Human Decision Processes*, vol. 151, p. 90–103, 2019.

[4] N. Castelo, M. W. Bos, and D. R. Lehmann, "Task-Dependent Algorithm Aversion," *Journal of Marketing Research*, vol. 56, no. 5, pp. 809–825, 2019.

[5] X. Luo, S. Tong, Z. Fang, and Z. Qu, "Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases," *Marketing Science*, 2019.

[6] S. Schanke, G. Burtch, and G. Ray, "Estimating the Impact of 'Humanizing' Customer Service Chatbots," jan 2020.

[7] M. Kosfeld, M. Heinrichs, P. J. Zak, U. Fischbacher, and E. Fehr, "Oxytocin increases trust in humans," *Nature*, vol. 435, no. 7042, pp. 673–676, 2005.

[8] C. Hildebrand, F. Efthymiou, F. Busquet, W. H. Hampton, D. L. Hoffman, and T. P. Novak, "Voice analytics in business research: Conceptual foundations, acoustic feature extraction, and applications," *Journal of Business Research*, vol. 121, no. January, pp. 364–374, 2020.

[9] N. Dahlbäck, Q. Wang, C. Nass, and J. Alwin, "Similarity is more important than expertise: Accent effects in speech interfaces," *Conference on Human Factors in Computing Systems - Proceedings*, pp. 1553–1556, 2007.

[10] C. Nass and K. M. Lee, "Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction, and consistency-attraction.," *Journal of Experimental Psychology: Applied*, vol. 7, no. 3, p. 171–181, 2001.

[11] Y. Jia, Y. Zhang, R. J. Weiss, Q. Wang, J. Shen, F. Ren, Z. Chen, P. Nguyen, R. Pang, I. L. Moreno, and Y. Wu, "Transfer learning from speaker verification to multispeaker text-to-speech synthesis," *Advances in Neural Information Processing Systems*, vol. 2018-December, no. NeurIPS, pp. 4480–4490, 2018.

[12] G. Charness and M. Dufwenberg, "Promises and partnership," *Econometrica*, vol. 74, no. 6, pp. 1579–1601, 2006.

[13] G. Charness and M. Dufwenberg, "Bare promises: An experiment," *Economics Letters*, vol. 107, no. 2, pp. 281–283, 2010.

[14] B. Libai, Y. Bart, S. Gensler, C. F. Hofacker, A. Kaplan, K. Kötterheinrich, and E. B. Kroll, "Brave new world? on ai and the management of customer relationships," *Journal of Interactive Marketing*, vol. 51, pp. 44–56, 2020.

[15] I. Benbasat and W. Wang, "Trust in and adoption of online recommendation agents," *J. AIS*, vol. 6, 03 2005.

[16] G. M. Becker, M. H. Degroot, and J. Marschak, "Measuring utility by a single-response sequential method," *Behavioral Science*, vol. 9, no. 3, p. 226–232, 1964.

[17] M. E. Schweitzer, J. C. Hershey, and E. T. Bradlow, "Promises and lies: Restoring violated trust," *Organizational Behavior and Human Decision Processes*, vol. 101, no. 1, pp. 1–19, 2006.

[18] X. Luo, M. S. Qin, Z. Fang, and Z. Qu, "Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions," *Journal of Marketing*, vol. 85, no. 2, pp. 14–32, 2021.

[19] M. T. Adam, T. Teubner, and H. Gimpel, "No Rage Against the Machine: How Computer Agents Mitigate Human Emotional Processes in Electronic Negotiations," *Group Decision and Negotiation*, vol. 27, no. 4, pp. 543–571, 2018.

[20] P. Tambe, P. Cappelli, and V. Yakubovich, "Artificial intelligence in human resources management: Challenges and a path forward," *California Management Review*, vol. 61, no. 4, p. 15–42, 2019.

[21] G. Adomavicius, J. C. Bockstedt, S. P. Curley, and J. Zhang, "Effects of online recommendations on consumers' willingness to pay," *Information Systems Research*, vol. 29, no. 1, pp. 84–102, 2018.

[22] A. Ansari, S. Essegaier, and R. Kohli, "Internet recommendation systems," *Journal of Marketing Research*, vol. 37, no. 3, pp. 363–375, 2000.

[23] K. Hosanagar, D. Fleder, D. Lee, and A. Buja, "Will the global village fracture into tribes recommender systems and their effects on consumer fragmentation," *Management Science*, vol. 60, no. 4, pp. 805–823, 2014.

[24] E. Brunswik, "Representative design and probabilistic theory in a functional psychology," *Psychological Review*, vol. 62, no. 3, pp. 193–217, 1955.

[25] R. M. Dawes, D. Faust, and P. E. Meehl, "Clinical Versus Actuarial Judgement," *Science*, vol. 243, no. 4899, p. 1668, 1989.

[26] N. Karelaia and R. M. Hogarth, "Determinants of Linear Judgment: A Meta-Analysis of Lens Model Studies," *Psychological Bulletin*, vol. 134, no. 3, pp. 404–426, 2008.

[27] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them," *Management Science*, vol. 64, no. 3, p. 1155–1170, 2018.

[28] M. Yeomans, A. Shah, S. Mullainathan, and J. Kleinberg, "Making sense of recommendations," *Journal of Behavioral Decision Making*, vol. 32, no. 4, pp. 403–414, 2019.

[29] J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan, "Human decisions and machine predictions*," *The Quarterly Journal of Economics*, 2017.

[30] C. F. Camerer, "Artifi cial Intelligence and Behavioral Economics," no. May, pp. 587–608, 2019.

[31] C. March, "The Behavioral Economics of Artificial Intelligence: Lessons from Experiments with Computer Players," *CESifo Working Paper*, no. 7926, 2019.

[32] R. H. Thaler, "Behavioral economics: Past, present, and future," *Revista de Economia Institucional*, vol. 20, no. 38, pp. 9–43, 2018.

[33] K. McCabe, D. Houser, L. Ryan, V. Smith, and T. Trouard, "A functional imaging study of cooperation in two-person reciprocal exchange," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 20, pp. 11832–11835, 2001.

[34] A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen, "The neural basis of economic decision-making in the ultimatum game," *Science*, vol. 300, no. 5626, pp. 1755–1758, 2003.

[35] S. Kiesler, L. Sproull, and K. Waters, "A prisoners dilemma experiment on cooperation with people and human-like computers.," *Journal of Personality and Social Psychology*, vol. 70, no. 1, p. 47–65, 1996.

[36] E. Torta, E. V. Dijk, P. A. M. Ruijten, and R. H. Cuijpers, "The ultimatum game as measurement tool for anthropomorphism in human–robot interaction," *Social Robotics Lecture Notes in Computer Science*, p. 209–217, 2013.

[37] M. K. Lee, S. Kiesler, and J. Forlizzi, "2011-CHI-behavioral_econ," pp. 325–334, 2011.

[38] L. Moretti and G. D. Pellegrino, "Disgust selectively modulates reciprocal fairness in economic interactions.," *Emotion*, vol. 10, no. 2, p. 169–180, 2010.

[39] J. Andreoni and J. Miller, "Rational Cooperation in Finitely Repeated Prisoner's Dilemma: Experimental Evidence," vol. 103, pp. 570–585, 1993.

[40] T. Teubner, M. Adam, and R. Riordan, "The impact of computerized agents on immediate emotions, overall arousal and bidding behavior in electronic auctions," *Journal of the Association for Information Systems*, vol. 16, no. 10, pp. 838–879, 2015.

[41] P. Duersch, A. Kolb, J. Oechssler, and B. C. Schipper, "Rage against the machines: How subjects play against learning algorithms," *Economic Theory*, vol. 43, no. 3, pp. 407–430, 2010.

[42] Y. Moon and C. Nass, "How "Real" Are Computer Personalities? Psychological Responses to Personality Types in Human-Computer Interaction," *Communication Research*, vol. 23, no. 6, pp. 651–674, 1996.

[43] D. Byrne, "Interpersonal attraction and attitude similarity.," *The Journal of Abnormal and Social Psychology*, vol. 62, no. 3, pp. 713–715, 1961.

[44] J. N. Bailenson, S. Iyengar, N. Yee, and N. A. Collins, "Facial similarity between voters and candidates causes influence," *Public Opinion Quarterly*, vol. 72, no. 5, pp. 935–961, 2008.

[45] P. G. Banikiotes and G. J. Neimeyer, "Construct importance and rating similarity as determinants of interpersonal attraction," *British Journal of Social Psychology*, vol. 20, pp. 259–263, 1981.

[46] C. Hu, K. M. Thomas, and C. E. Lance, "Intentions to initiate mentoring relationships: Understanding the impact of race, proactivity, feelings of deprivation, and relationship roles," *Journal of Social Psychology*, vol. 148, no. 6, pp. 727–744, 2008.

[47] R. Singh, Y. Y. Tay, and K. Sankaran, "Causal role of trust in interpersonal attraction from attitude similarity," *Journal of Social and Personal Relationships*, vol. 34, no. 5, pp. 717–731, 2017.

[48] D. T. Yeong Tan and R. Singh, "Attitudes and Attraction: A Developmental Study of the Similarity-Attraction and Dissimilarity-Repulsion Hypotheses," *Personality and Social Psychology Bulletin*, vol. 21, pp. 975–986, sep 1995.

[49] P. Adamopoulos, A. Ghose, and V. Todri, "The impact of user personality traits on word of mouth: Text-mining social media platforms," *Information Systems Research*, vol. 29, no. 3, pp. 612–640, 2018.

[50] S. Aral, L. Muchnik, and A. Sundararajan, "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 51, pp. 21544–21549, 2009.

[51] M. Kaptein, D. Castaneda, N. Fernandez, and C. Nass, "Extending the similarity-attraction effect: The effects of when-similarity in computer-mediated communication," *Journal of Computer-Mediated Communication*, vol. 19, no. 3, pp. 342–357, 2014.

[52] T. Trainer, J. R. Taylor, and C. J. Stanton, *Choosing the Best Robot for the Job: Affinity Bias in Human-Robot Interaction*, vol. 12483 LNAI. Springer International Publishing, 2020.

[53] D. Byrne and G. L. Clore, "A reinforcement model of evaluative responses.," 1970.

[54] E. Berscheid and E. Hatfield, *Interpersonal attraction*. Reading, Mass.: Reading, Mass. : Addison-Wesley Pub. Co., 2d ed.. ed., 1978.

[55] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[56] K. Imai, L. Keele, and T. Yamamoto, "Identification, inference and sensitivity analysis for causal mediation effects," *Statistical science*, pp. 51–71, 2010.

[57] J. B. Walther, "Interpersonal effects in computer-mediated interaction," *Communication Research*, vol. 19, no. 1, p. 52–90, 1992.

[58] J. B. Walther and L. C. Tidwell, "Nonverbal cues in computer-mediated communication, and the effect of chronemics on relational communication," *Journal of Organizational Computing*, vol. 5, no. 4, p. 355–378, 1995.

[59] T. Sun, Z. Yuan, C. Li, K. Zhang, and J. Xu, "The value of personal data in internet commerce: A high-stake field experiment on data regulation policy," *SSRN Electronic Journal*, 2020.