Guess Who? An Empirical Study of Gender Deception and Detection in Computer-Mediated Communication

Shuyuan Mary Ho
Florida State University
School of Library and Information Studies
142 Collegiate Loop
Tallahassee, FL 32306-2100
smho@fsu.edu

Jonathan M. Hollister
Florida State University
School of Library and Information Studies
142 Collegiate Loop
Tallahassee, FL 32306-2100
jmh09k@my.fsu.edu

ABSTRACT
The verification of an online conversation partner's identity is a challenge due to the lack of verbal and visual cues in computer-mediated communication. People must constantly assess the identity of whomever they are communicating with based on limited interaction. This poster describes an empirical study that identifies how people attribute gender and detects gender deception in online text-based communication. Data has been collected through an interactive social media game platform and designed as a quasi-experiment. Our study confirms that the speakers' domain knowledge, perceived trustworthiness, and self-efficacy have an impact on the success of gender deception and attribution. Perceived gender of the speaker also had an impact on the successful detection of gender deceptive messages.

Keywords
Human information behavior, gender, deception, self-efficacy, trustworthiness, online game, computer-mediated communication.

INTRODUCTION
Today's society relies on text-based computer-mediated communications (CMC) systems to interact and exchange information. The messages and information transferred in cyberspace influences the decisions and actions a user might take. The user, or the recipient, of information must take a risk by putting trust in their communication partners. However, without the synchronous visual cues, a user may face difficulties in assessing the truthfulness of the identity attributes (e.g., gender, etc.) of those whom they are communicating with. Thus, online information user is vulnerable to online deception (Hancock 2007).

Our goal in this study is to better understand the factors that enable successful gender imitation and gender attribution in online communication by exploring the research question:

> *What factors enable gender deception and detection in CMC environments*?

ASIST 2013, November 1-6, 2013, Montreal, Quebec, Canada.

The following discussions will cover our theoretical foundations with several hypotheses. The method section describes the design of this empirical study. Preliminary results with statistical evidence are discussed in support of our hypotheses.

THEORETICAL FOUNDATIONS
This study attempts to integrate factors that have been associated with trust, deception outcome, gender, and self-efficacy in order to understand the factors of detecting gender deception in online environments.

Gender Deception
Gender imitation or misrepresentation is a significant aspect of online deception. Herring and Martinson (2004) suggest that the language of those intentionally trying to deceive others online about their gender still use stylistic features characteristic of one's actual gender. To facilitate online fraud activity, CMC users may attempt to imitate the opposite gender in order to deceive their conversation partner. As online interaction and ecommerce progress to become a norm, users face increasing difficulties in determining gender in online communication. Successful identity thieves can take advantage of a user's inability to detect gender. Thus, we hypothesize that:

- H1a: Gender of the speaker is associated with the success of deception of gender deceptive messages.
- H1b: Perceived gender of the speaker is associated with the success of detection of gender deceptive messages.

Trust
Trust involves an expectation of "ethically justifiable behavior – that is, morally correct decisions and actions based upon ethical principles of analysis" (Ho 2009). It necessitates a relationship of reliance between a speaker (e.g., deceiver) and recipient (e.g., detector) of text messages in an online environment.

Trustworthiness
Ho (2009) develops a theory of trustworthiness attribution for identifying potential insider threats in cyber organizations. These cyber threats are situations initiated by human betrayal; a cognitive state that signifies the violation and defilement of interpersonal trust (Elangovan et al. 1998). Ho (2009) suggests that a negative impression of a speaker's trustworthiness can be observed

and interpreted through communications with a recipient who has an established relationship with the speaker, and posits that a speaker's trustworthiness can be accurately attributed with internal causality by the recipient.

However, attributing trustworthiness and detecting deception involve multiple factors. Interpersonal deception theory suggests that both speakers and recipients of deceptive messages will dynamically alter and utilize strategic communicative behaviors in order to deceive (and detect) deceptive intent (Buller et al. 1996).

Deception is a "message knowingly transmitted by a speaker to foster a false belief or conclusion by the receiver" (Buller et al. 1996). A speaker may intentionally manipulate recipients with deceptive messages. Perceived trustworthiness, the expectance that the trustee's actions are consistent with the outcomes observed by the trustor over time, may provide clues that uncover a speakers' intent to deceive (Ho 2009). We thus hypothesize that:

- H2: Perceived trustworthiness of the speaker is positively associated with the recipient's ability to detect correct gender.

Trustfulness

On the other side of the spectrum, the trustfulness, the predisposition to trust others, of the recipient is also a factor in detecting deception. A more trustful receiver may irrationally and dangerously attribute high trustworthiness of a speaker due to lack of cognitive ability, domain knowledge, prior relations or emotional connections with the speaker (McAllister 1995). Thus, we hypothesize that:

- H3: Trustfulness of the recipient is negatively associated with their ability to detect deceptive messages.

Domain Knowledge & Self-Efficacy

Deceptions, at their base, are behaviors that are affected by self-efficacy, an individual's beliefs concerning their capability to perform a certain task (Bandura 1977). A speaker may be motivated to deceive, but lack the confidence in their ability to successfully accomplish it; this belief may affect the level of effort or even the decision to attempt deception. Thus, we hypothesize that:

- H4: Domain knowledge is positively associated with the speaker's ability to send successful deceptive message.
- H5: Speakers with higher self-efficacy can deceive better.

METHOD

We designed an asynchronous interactive Facebook game entitled "*Guess Who?*," as an instrument to collect data. The study was reviewed and approved by Florida State University Institutional Review Board (IRB) in September 2012. Participants' demographic information, especially the ground truth information of their gender, were authenticated and verified by Facebook authentication mechanisms. Participants were recruited from November 2012 through January 2013 via social media marketing on Facebook and Twitter, and listserv emails from students at Florida State University.

Data Collection

Participants were welcomed into the *Virtual Funhouse*, a socio-technical game portal, to play the game after obtaining their consent for participation. This research platform protects participants' rights to privacy and confidentiality and is physically located in-house at Florida State University College of Communication and Information. The total participants of this study were 26, with 11 females (approx. 42.3%) and 15 males (approx. 57.7%). These 26 participants generated 98 records (n=98) suitable for analysis.

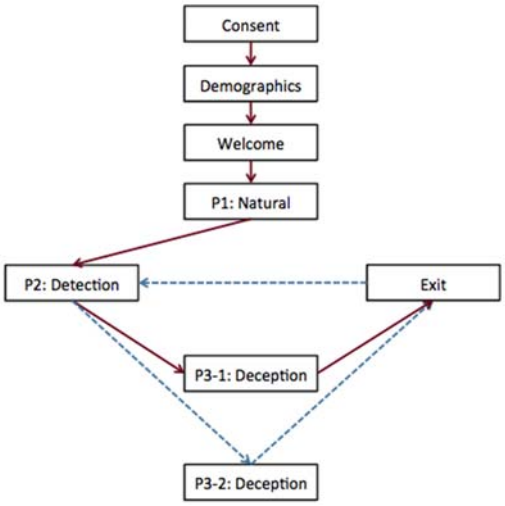

Figure 1: Game Design Schematics

Game Design

The participants entered into three phases of the game (Figure 1): natural (truth-telling), detection, and deception. In natural (truth-telling) phase (P1), participants selected one most knowledgeable topic and one lesser-known topic from a list of topics that included popular hobbies, sports, and current events. They wrote a truthful statement for each selected topic. Then, we collected each participant's self-assessment as to the level of domain knowledge for each statement.

During the detection phase (P2), participants were asked to attribute gender and evaluate the perceived level of domain knowledge, and trustworthiness of a speaker based on topic descriptions randomly selected from other participants. Participants were asked to assess their domain knowledge regarding the reviewed statement of a particular topic.

In the deception phase (P3-1/P3-2), participants were asked to provide a statement as if they were of the opposite gender, based on their selected topics from P1. Prior to exiting the game, participants were asked about their confidence level when attributing gender, as well as imitating the opposite gender. Levels of participants' motivation and self-efficacy during the overall game were also captured at this stage.

RESULTS AND DISCUSSION

Each of the hypotheses were tested and compared across overall, truthful state, and deceptive state groups using SPSS. Below we discuss the results of our data analysis, in support of the hypotheses testing.

Deception Outcome & Gender: H1a Not Supported

Our data confirms prior research that people are poor lie detectors (Buller et al. 1996). In the overall sample, 46.9% of the participants were successful in gender detection (n=98, SD=.502). There were 47.4% (n=38, SD=.506) successfully attributing gender during the natural (truth-telling) phase, and 46.7% (n=60, SD=.503) successfully attributing gender during the deceptive phase. There was no significant difference in successful gender deception or detection between females and males in any group.

Perceived Gender: H1b Partially Supported

In the overall sample, a Pearson Chi-Square test found statistical significance between the perceived gender of the speaker and the success outcome of the deceptive message ($\chi 2$=7.148, df=1, p<.01, n=98). Similarly, there was also significance in the truthful state group ($\chi 2$=5.546, df=1, p<.05, n=38); but, there was no Chi-Square significance in the deceptive state group. In the overall and natural groups, receivers that perceived the speakers to be female were successful in their attributions 59.6% and 63.6% of the time, respectively, as compared to 32.6% and 25%, respectively, success rates of receivers that perceived the speakers to be male. These findings suggest that females are more easily detectable. However, since there was no significant difference in the deceptive state group, deceptive priming may make it more difficult for the receiver to correctly identify gender.

In the overall sample, speakers perceived to be men were significantly (t=2.239, df=96, p<.05) perceived as less trustworthy (m=2.7391, SD=.72064, n=46) than those who perceived as women (m=3.0753, SD=.76024, n=52).

Trustworthiness: H2 Supported

Our data supports hypothesis 2 that perceived trustworthiness of the speaker (index of 12 items with Cronbach's alpha=.898) is positively associated with the recipient's ability to detect correct gender. There was significant difference (t=3.278, df=20.309, p<.01, equal variances not assumed) within the truth-telling state group. Recipients that perceived speakers more trustworthy were more successful in gender detection (m=.59, SD=.501, n=29) and receivers that perceived speakers less trustworthy were less successful in gender detection (m=.11, SD=.333, n=9).

A t-test of the overall sample found significance (t=9.176, df=44.44, p<.01, equal variances not assumed) between the perceived trustworthiness and perceptions of high or low domain knowledge. Recipients that perceived speaker to have high levels of domain knowledge also found them more trustworthy (m=3.7231, SD=.64420, n=31). Similarly, receivers who perceived the speaker as having low levels of domain knowledge also found them to be less trustworthy (m=2.5448, SD=.45605, n=67).

Trustfulness: H3 Not Supported

The index of utilized for trustfulness (α<.60), consisting of 8 items, did not have an acceptable reliability rating. As such, the authors cannot suitably test or support H3.

Domain Knowledge: H4 Partially Supported

When domain knowledge rating levels (scale 1-10) were grouped into high ($\geq$6) and low (<6) levels, a t-test found significance (t=2.732, df=18.275, p<0.5, equal variances not assumed) between the means of successful gender detection in the natural group; which were 0.80 (n=10, SD=0.422) and 0.36 (n=28, SD=0.488), respectively. In the deceptive state, a t-test found significance (t=-2.047, df=44, p<0.5) between the means of successful gender detection in the deceptive group; which were 0.37 (n=19, SD=0.496) and 0.67 (n=27, SD=0.480), respectively. There was no significant difference in success rate between high and low domain knowledge participants in the overall sample.

Within the natural state group, a t-test found significance (t=-2.095, df=27.002, p<.05, equal variances not assumed) in the difference between the means of domain knowledge rating of the speaker when compared to the success or failure of the detector to attribute gender. The domain knowledge of the speaker was correlated 0.337 (p<.05) with the success outcome in the natural state. Successful gender deceptive speakers had significantly lower average domain knowledge ratings (m=2.55, n=20, SD=2.328) than the unsuccessful speakers (m=4.78, n=18, SD=3.934), respectively.

Within the deceptive state group, a t-test found a significant difference in the means of domain knowledge ratings when compared to outcome (t=2.713, df=55, p<.01). The t-test mean for domain knowledge rating of the speakers of successfully gender deceptive message speakers (m=6.17, n=29, SD=3.485) was significantly higher than the mean for speakers that were unsuccessful (m=3.82, n=28, SD=3.031). The Pearson correlation for this association was also significant -.344 (p<.01).

These results suggest imply that when speakers are at the natural (truth-telling) state, the less domain knowledge they have, the greater they are at confusing the opposite

gender. However, in the deceptive state, successfully gender imitating speakers have significantly higher domain knowledge ratings (t=3.278, df=20.309, p<.001, equal variances not assumed. This suggests that the mechanics in the game, gender deception priming, may have influenced player's ability and performance to attribute and imitate gender.

Self-efficacy: H5 Supported

A t-test found significant differences in mean self-efficacy (scale 1-7) between outcomes of gender deceptive messages (t=2.161, df=47, p<.05). Successful gender deceptive message speakers were more confident (m=4.92, n=24, SD=1.792) that their messages would be successful at gender deception than unsuccessful speakers (m=3.88, SD=1.563, n=25). Also, a t-test found significant difference (t=4.115, df=44, SD=.45, p<.01) between the mean self-efficacy levels for those with high domain knowledge (≥6) and for speakers with lower domain knowledge (<6); the means were 5.37 (n=19, SD=1.739) and 3.52 (n=27, SD=1.312), respectively.

LIMITATIONS

The online game environment in this study shares the limitations of other quasi-experimental methods. For example, the increased control of the setting threatens contextual realism. The small sample size also affects the generalizability of the findings. However, random selections of our research participants, and the randomized display of either truthful or deceptive statements to those in the detection phase, enable the results to be generalizable.

There is a slightly unequal gender distribution within the sample—42.3% female and 57.7% male—but this difference is statistically insignificant when using a z-test (z=-1.1094, p>.05). Given the findings of past research in this area and considering chance, it is possible that a detector could select male every time that s/he would play, and have better than 50% gender attribution correctness. Gender bias may also be inherent in the list of topics available for participants to discuss; however, allowing participants to self-select the topics they discuss may reduce this bias. Additionally, our study design did not incorporate considerations of nonbinary genders.

As mentioned above, the trustfulness index did not reach a suitable reliability level (>.60) even after removing items with low correlations. The design of the game, specifically the gender role-playing and multi-phased structure, may have influenced the participants' trustfulness of others. Future studies will need to include more precise measures to understand whether trustfulness impacts gender deception and attribution.

The historical and ongoing relations between social actors are important factors in making attributions regarding truthful or deceptive communication. Due to the asynchronous and randomized nature of this online game, the players did not know each other and were not given the ability to display longitudinal relationship dynamics, such as consistency. As a result, players lacked knowledge of these historical and relational aspects.

CONCLUSIONS AND FUTURE RESEARCH

As more people conduct online business, utilize e-government, and socialize in CMC environments, the threat of online deception increases. In order for online users to maintain trusting relationships with other individuals and institutions, online user's ability to detect gender deception based on limited information behavior becomes imperatively important. Our findings will assist both practitioners and researchers in understanding the role of gender, as an attribute of identity, in trust and deception within CMC environments. Our continued efforts in this research stream will focus on collecting larger samples empirically tested with a path analysis. This model will quantify hypothesized relationships among observed variables and factors. We will also additionally study the specific strategic and nonstrategic language cues that both detectors and intentional deceivers deploy.

ACKNOWLEDGMENTS

The first author thanks the National Science Foundation for the support of the Secure and Trustworthy Cyberspace EAGER award #1347113 09/01/13-08/31/15.

REFERENCES

Bandura, A. 1977. "Self-efficacy: Toward a unifying theory of behavioral change," *Psychological Review* (84:2), pp 191-215.

Buller, D. B., and Burgoon, J. K. 1996. "Interpersonal deception theory," *Communication Theory* (6:3), pp 203-242.

Elangovan, A. R., and Shapiro, D. L. 1998. "Betrayal of trust in organizations," *Academy of Management* (23:3), pp 547-566.

Hancock, J. T. 2007. "Digital deception: When, where and how people lie online," in *Oxford Hanbook of Internet Psychology,* K. McKenna, T. Postmes, U. Reips and A. N. Joinson (eds.), Oxford University Press: Oxford.

Herring, S. S., and Martinson, A. 2004. "Assessing gender authenticity in computer-mediated language use: Evidence from an identity game," *Journal of Language and Social Psychology* (23:4), pp 424-446.

Ho, S. M. 2009. *Behavioral anomaly detection: A socio-technical study of trustworthiness in virtual organizations,* Information Systems, Syracuse University, Syracuse.

McAllister, D. J. 1995. "Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations," *The Academy of Management Journal* (38:1) February, pp 24-59.