ASSESSING GENDER AUTHENTICITY IN COMPUTER-MEDIATED LANGUAGE USE
Evidence From an Identity Game

SUSAN C. HERRING
ANNA MARTINSON
Indiana University

Although a substantial body of research exists on gender differences in computer-mediated communication, relatively little empirical attention has been directed toward how people perform a different gender online, or to what behavioral cues other participants attend in assessing others' real-life gender. This study analyzes deceptive gender performances and assessments of their authenticity in The Turing Game, a publicly available synchronous text chat environment that supports spontaneous identity games. Content analysis of game logs shows that contestants produce stereotypical content when attempting to pass as the opposite gender, as well as giving off stylistic cues to their real-life gender. However, contrary to previous evidence that people judge online gender authenticity on the basis of linguistic styles, the judges in The Turing Game base their assessments mostly on stereotyped content, leading to a high rate of error. These findings are interpreted in terms of signal costs and conscious accessibility of cues.

Keywords: *computer-mediated communication (CMC); deception; discourse style; gender; identity; stereotypes*

The Internet, it was said, would mask the visible differences between men and women, between races, and across other social categories, because e-mail and other forms of computer-mediated communication (CMC) strip away physical appearance cues, leaving us to express ourselves online only through language. Yet, signaling gender is often an activity of interest online, for a variety of reasons, and ascertaining the gender of others is particularly important to users who for one reason or another doubt the sincerity of their online conversation

---

partners. More generally, it is important to know an interlocutor's identity in order to understand and evaluate the interaction; this is especially true for gender, which is conventionally associated with different norms, roles, and communication styles in most human cultures. How can one determine the gender of participants in public Internet communication spaces when language provides the only cues?

This question has troubled and fascinated CMC users and scholars alike since the early days of the Internet, giving rise to debates about "anonymity" and warnings about the social risks inherent in online communication, particularly in the context of online relationship formation (Bell & de la Rue, 1995; Van Gelder, 1985). Individuals may nominally indicate that they are female or male, but identity deception is easier to carry out in text-based computer-mediated communication than face-to-face: CMC allows participants to control their manner of self-presentation through selective posting of typed information, and the Internet typically attracts strangers in different geographical locations who are unlikely ever to meet face-to-face.

A substantial body of research exists on gender differences in CMC. Its canon includes theoretical claims about the "performed" nature of gender identities, and documentation of cases of gender play. However, relatively little empirical attention has been directed toward how people perform a different gender online, or to what behavioral cues other participants attend in assessing others' real-life gender. What little research has been conducted on these questions has tended to be experimental, a reflection of the difficulty of ascertaining people's real-life gender, or what others think their real-life gender is, in natural online interaction.[1] There is also inconsistency within the empirical research on gender-linked language in CMC with regard to how the real-life gender of users has been authenticated, further underscoring the need for research examining the relationship among claimed gender, actual gender, and language use.

In this study, we analyze deceptive gender performances and the assessments of their authenticity that occurred in an online context known as The Turing Game. Publicly available via the Internet, The Turing Game features a text-based chat environment that supported spontaneous, real-time communication for the purpose of "To Tell the Truth"-style identity games,[2] the most popular of which are games about gender identity. We investigate the stylistic and content choices of the contestants in relation to their claimed and revealed genders, comparing these with judges' ratings of who is really male or female. The results show that contestants produce stereotypical content when attempting to pass as the opposite gender, as well as persisting in giving off stylistic cues at the word and sentence levels to their real-life gender. However, contrary to previous evidence that people judge online gender authenticity on the basis of linguistic styles (Savicki, Kelley, & Oesterreich, 1999; Thomson & Murachver, 2001), the judges

in The Turing Game appear to base their assessments mostly on ste-
reotyped content, leading to a high rate of error in their identification
of contestants' gender. We interpret these findings in terms of signal
costs and conscious accessibility of cues, arguing that in this context,
stylistic gender differences, although more reliable signals, are not as
accessible as conventional beliefs about gender.

BACKGROUND

GENDER AND CMC

The notion that gender is a performance—something that one does,
rather than something one is—can be traced back to Birdwhistell's
(1964, 1970) concept of masculinity and femininity as "display." More
recently, it has been associated with the writings of feminist theorist
Judith Butler (1990), who "understand[s] gender as a relation among
socially constituted subjects in specifiable contexts" (p. 25). A number
of cyber-theorists have embraced this view, pointing out that because
text-based CMC filters out cues such as voice and physical appearance
(cf. Kiesler, Siegel, & McGuire, 1984), online communicators are theo-
retically free to perform any identity, gendered or otherwise, that they
can imagine (Danet, 1998; McRae, 1996; Stone, 1996). One version of
this position holds that online personae can evolve existences inde-
pendent from their flesh-and-blood animators, shaped by the social
interactions and the contexts they encounter online (Stone, 1996).

Observations of recreational Internet discussion forums provide
anecdotal evidence that gender can be successfully masked online.
Bruckman (1993) interviewed male and female participants in social
MOOs (text-based virtual reality environments) who report "gender
swapping"—presenting themselves online as a different gender—and
fooling others for months at a time. McRae (1996) interviewed MOO
participants who reported performing not only alternative genders but
alternative sexual orientations and even alternative species identities
and engaging in extended sexual and romantic encounters in these
assumed identities. Hall's (1996) description of a woman who boasts of
how easy it is to pretend to be a gay male online is one of the few indi-
cators of how such performances are actually carried off, in this case,
through adopting the stylistic features stereotypically associated with
gay male (in her words, "gay mail") language:

> I guess with e-mail we are what we write and I could *easily* imperson-
> ate a gay mail. "Hi hon. How are you today? I saw so-and-so and sister did
> he look *bad*. A *serious* fasion [sic] no-no. Throw *that* boy back to the
> straights" etc. (Hall, p. 153)

Not all online gender performances are convincing. A near-truism in recreational chat rooms is that if a "female" participant behaves in ways that seem too sexy to be true (e.g., by describing herself as "hot" or voluptuous, or aggressively offering or requesting sexual favors from male participants), it is probably a male pretending to be a female (Bruckman, 1993; Herring, 1998). However, aggression and sexual pursuit (of females) can be effective means for females to perform male identities in recreational chat rooms (Herring, 1998). That people can be fooled by deceptive gender performances in text-based CMC supports Walther's (1996) claim that online communicators tend to over-attribute characteristics of their interlocutors from minimal cues. In the case of gender performances, it seems that they fill in the gaps according to cultural scripts.

Whereas online gender deception is possible, other researchers argue that it is difficult to achieve. Pavel Curtis, the creator of the LambdaMOO environment, after observing interactions in the MOO over a period of several years, noted that pretending to be something one was not was relatively uncommon (Curtis, 1992; see also Roberts & Parks, 1999). Although people might take on a nickname of the opposite gender, for example, it was rare for them to maintain behaviors consistent with that gender over time, due to the strain of having to maintain an artificial persona.

Other scholars have amassed evidence that males and females tend to use language in different ways, presumably unconsciously, online as well as offline (Hall, 1996; Herring, 1993, 1994, 1996, 1998, 2003; Hills, 2000; Kendall, 1998; Thomson & Murachver, 2001): Males make greater use of assertions, self-promotion, rhetorical questions, profanity, sexual references, sarcasm, challenges, and insults, whereas females make greater use of hedges, justifications, expressions of emotion, representations of smiling and laughter, personal pronouns, and supportive and polite language. Unconscious use of gendered discourse styles can reveal one's actual gender even when one is performing a different gender (or trying not to give off any gender cues). Herring (1996) cites examples of individuals attempting to pass as a different gender in asynchronous discussion forums (what Hall [1996] calls "cross-expressing") who were outed as imposters on the basis of inconsistencies between their performed gender and their discourse style. Elsewhere, Herring (1998) reports male-to-female performances in Internet Relay Chat that failed to convince other participants because the individuals retained aggressive, face-threatening behavior and sexual coarseness as part of their performances. In the latter study, cross-expressing accounted for fewer than 5% of all gender-indexing behaviors in sample logs from six chatrooms.

Because gender differences in online discourse exist, and because they tend to map on to participants' real-life gender (e.g., Herring, 1992, 1996), it is often assumed that CMC users attend to discourse dif-

ferences in deciding who is female or male. Thomson and Murachver (2001) tested this assumption empirically by asking undergraduates to identify the gender of the authors of sample e-mails selected to illustrate features of gender-preferential styles. The female messages were correctly identified as female, even when indications of topic were removed from the sample. Savicki et al. (1999) also found evidence that readers can accurately predict the gender of authors of anonymous e-mail messages on the basis of stylistic features. Using discriminant analysis, Hills (2000) and Thomson and Murachver (2001) show that it is a combination of weighted features (e.g., adjectives, apologies, compliments, insults, intensive adverbs, modals, opinions, references to emotion, self-derogatory comments, subordinating conjunctions), rather than any single feature alone, that allows for accurate gender attribution, lending empirical support to the notion of gender "styles."

Few studies have investigated empirically what takes place when individuals consciously seek to deceive others about their gender online. An exception is Hills (2000), who instructed members of same-sex dyads to exchange e-mails with the intention of persuading their "netpal," an experimental partner with whom they were previously unacquainted, that they were a different gender, albeit without mentioning gender or alluding to it (e.g., by mentioning "my boyfriend/girfriend"). Although both male and female participants in the study rated their own cross-gender performances high in authenticity, their netpals did not rate their performances as very convincing (males gave a mean rating of 3.80 and females gave a mean rating of 4.33 on a scale where 1 was "definitely female" and 7 was "definitely male"). Hills attributes this to the tendency of participants to perform gender through the choice of (sometimes over-) stereotypical content, whereas their use of stylistic features at the word and sentence level remained largely true to their real-life genders. Similarly, Berman and Bruckman (2001) report that experienced Turing Game players are suspicious of overly stereotypical answers, rating those who produce them as less authentic, although as one female interviewee commented, "[G]ender is mostly about the stereotypical things, about how I react to clothes and men and things. And so the game is too. It should be, I guess I mean" (n.p.). When interviewed by Berman and Bruckman, a number of Turing Game players also expressed the belief that males produce shorter messages than do females, and make use of more sentence fragments, as though they were "too busy to use grammar."

Underlying the tension between stereotyped content and linguistic style are issues of cost of production of, and conscious accessibility to, gender identity cues. With regard to production, Donath (1999) draws a theoretical distinction between identity cues that are costly to produce and hence hard to fake (e.g., physical strength)—what she calls assessment signals—and cues that are less costly to produce, but open

to deception (e.g., wearing a bodybuilder T-shirt)—what she calls conventional signals. On one reading, all online manifestations of gender are conventional signals, in that they involve what people say, rather than how they look or what they can do (hence the widespread claim that gender is easy to fake online). At the same time, not all gender identity cues are equally accessible to conscious reflection and control: Some have attained the status of popular stereotypes (e.g., that women talk more than men), whereas others operate at a more abstract symbolic level (e.g., use of personal pronouns, mitigation, rhetorical questions, and their clustering into gender styles). To the extent that discourse styles are harder to access and manipulate consciously—in other words, are more costly to produce—than more conventional manifestations of online gender, they might be said to function as assessment signals and thus, to be more reliable indicators of gender authenticity. This theoretical proposition is examined empirically in the investigation described below.

THE TURING GAME

The Turing Game was developed by Joshua Berman and Amy Bruckman as a research project of the Electronic Learning Communities group at The Georgia Institute of Technology, and released on the Internet as a free downloadable chat client in July 1999. The name pays homage to Alan Turing, the British mathematician who pioneered principles leading to modern computing, and who was responsible for the concept of the Turing Test, "an 'imitation game', in which a human being and a computer would be interrogated under conditions where the interrogator would not know which was which, the communication being entirely by textual messages,"[3] and that might thus phenomenologically erase the difference between computers and humans.[4]

The designers describe their purpose in creating the environment as follows:

> We have created a game to help us understand issues of online identity. In this environment, which we call The Turing Game, a panel of users all pretend to be a member of some group, such as women. Some of the users, who are women, are trying to prove that fact to their audience. Others are men, trying to masquerade as women. An audience of both genders tries to discover whom the imposters are, by asking questions and analyzing the panel members' answers. Games can cover aspects of gender, race, or any other cultural marker of the users' choice. (From The Turing Game Web site, http://www.cc.gatech.edu/elc/turing/info2_5.html)

In this article, in place of the somewhat ambiguous terms *users* (or *panel members*) and *audience*, we employ the terms *contestants* and *judges*, as in the television game show that The Turing Game resem-

bles, with the caveat that these roles are assigned ad hoc, rather than fixed. That is, a moderator is appointed for each game who selects one or more contestants from the interested parties present.[5] The remaining participants become judges and ask questions of the contestants directly or through the moderator. A participant who is a judge in one game may become a contestant in the next game, and vice versa. Moderators may also become contestants or judges, although only trusted regulars appear to serve as moderators. At the end of the game play, all players participate in a debriefing chat session in which the contestants reveal their real-life gender.

Within one year after its release, The Turing Game had been played by 11,158 people around the world. Of the 2,212 games played, 973 (44%) were games about gender identity (Berman & Bruckman, 2001). For gender games, contestants are first instructed to select a nickname "that you think a man/woman would choose," and then to "introduce yourself." Berman and Bruckman (2001) report the start of a typical male game, as follows:

Brian: Hello
Allan: I'm 6'3" Blk hr, 220 lbs
Warren: Hello, Go Seminoles

In this example, Brian and Allan are females in real life, and Warren is a real-life male. Each judge has a sliding scale from 0 to 10 on his or her interface set at 5.0 at the beginning of each game that she or he can adjust as the game proceeds, and from which a composite score for each contestant is derived and displayed at the end of the game. A score of "0" means an unconvincing gender performance (in the male game example above, that the contestant is definitely female), and a score of "10" means a convincing performance (e.g., in a male game, that the contestant is definitely male). After the scores are displayed, the participants exit the game space and engage in an informal debriefing chat in which the contestants are required by the rules of the game to reveal their real-life genders. All keystrokes made during the game are automatically recorded, and log files of the game and the debriefing chat are posted to an archive on the game's Web site, along with the composite ratings for each contestant, to promote reflection on the performances (Berman & Bruckman, 2001).

HYPOTHESES

Hypothesis 1 (H1): Contestants will communicate in ways appropriate to the gender they are performing.

In other words, we expect that contestants of both genders and in both types of gender games will attempt to persuade the judges

through their communication that they are really the gender they claim to be. Specifically, we hypothesize the following:

H1a: Contestants will choose nicknames appropriate to the gender they are performing.

H1b: Contestants will make use of stylistic features (hedges, boosters, apologies, profanity, etc.) appropriate to the gender they are performing.

H1c: Contestants will produce topical content appropriate to the gender they are performing (that is, they will talk about topics males or females are conventionally expected to talk about, and give content-appropriate answers to questions asked by the judges and moderator).

At the same time, in keeping with previous research findings on the persistence of real-life gender cues in online communication, we hypothesize that

H2: Contestants will preserve features of their real-life gender, regardless of the gender they are performing.

It is possible for both H1 and H2 to be true to varying degrees; this is the result we expect to obtain.

We further hypothesize that there will be a relationship between contestants' communicative behaviors and the ratings contestants receive from judges, namely:

H3: Contestants who communicate in ways appropriate to the gender they are performing will be rated as more authentic.

Specifically,

H3a: Contestants who choose nicknames appropriate to the gender they are performing will be rated as more authentic.

H3b: Contestants who make use of stylistic features (hedges, boosters, apologies, profanity, etc.) appropriate to the gender they are performing will be rated as more authentic.

H3c: Contestants who produce topical content appropriate to the gender they are performing will be rated as more authentic.

It follows from the intersection of H2 and H3 that

H4: Contestants performing their real-life gender will be rated as more authentic than those performing a different gender.

That is, because contestants will enact stereotypical gender-linked language, and add to it unintended but gender-consistent language behavior, there will be less room for error in real-gender performances than in cross-gender performances.

RESEARCH QUESTIONS

Based on the hypotheses articulated above, this study asks three general research questions (RQs):

RQ1: How do contestants in gender identity games present themselves? (i.e., what nicknames do they choose, what stylistic features do they employ, what content do they produce in response to the judges' questions?) Are there differences between real-life males and real-life females, between same-sex or cross-sex performances, and/or between male and female identity games?

RQ2: To what aspect(s) of contestants' self-presentation do judges attend when assessing gender authenticity? Which aspects are most important in judges' decisions?

RQ3: How successful are contestants' self-presentation strategies and judges' assessment strategies in terms of their respective goals (i.e., convincing the judges that one is the gender one claims to be, or discerning the actual gender of the contestants)?

METHOD

DATA

The raw data for this study are the publicly available synchronous log files of game play and debriefing chats, as well as the composite ratings for each contestant, for 39 gender games played between August 1999 and February 2000, the period during which The Turing Game was most actively played. These sources provide three key types of information: the contestants' gender performances (game logs), the judges' assessment of the authenticity of the performances (ratings), and the contestants' real-life genders (debriefing chats). We sampled the first 15 gender games per month for every other month during the active period (August, October, December, and February), for a total of 50 games,[6] of which 39 were usable for analysis.[7] Of these, 19 were female-presenting games, and 20 were male-presenting games.[8] In all, 73 contestants are included in the sample, of whom 38 were real-life females, and 35 were real-life males. The game play (questions and answers) averaged 41.6 messages per game, for a total of 1,622 messages that were subjected to analysis.

ANALYTICAL METHODS

Content Analysis

Content analysis was employed to analyze three types of information contained in the game files: nickname, stylistic features, and topi-

cal content. The authors coded nicknames; one of the authors plus an additional coder categorized stylistic features and topical content. Interrater agreement rate was in excess of 80% for all three types. Coding disagreements were resolved through discussion.

Nicknames were coded as female, male, or neutral on the basis of their explicit or implicit gender associations. Examples from the corpus of nicknames of each type are given below.

Nickname Examples

Female	Eve, RiotGrrl, Sweetie
Male	Alex, BillieBoy, Lost Viking
Neutral	Chris, serfer, Number 3

In the above examples, the first item in each row is a traditional female, male, or gender-neutral first name; the second item contains a human descriptor that in the case of the first two rows is gendered ("Grrl," "Boy"); and in the third row is gender neutral ("serfer").

For the analysis of stylistic features, individual features identified in previous research on gender differences in CMC (e.g., Herring, 1993, 1994, 1995, 1996, 1998; cf. Lakoff, 1975) were coded and counted in the game play, then combined into two gender-preferential styles (see Hills, 2000, who argues that composite styles are more indicative of gender identity than individual features). The features that were hypothesized to be characteristic of female- and male-preferential styles are listed below:

Female stylistic features included *hedges* (just, sort of, a bit, quite, etc.), *adverbial qualifiers* (maybe, possibly, probably), *possibility modals* (could, might, may), *politeness* (thanks, apologies, "please"), *hesitation fillers* (um, er), *emoticons and/or laughter* (:-), ha ha), *evidentials* (I guess, I think, It seems), and *clausal mitigation* ("This sounds dumb, but . . . ").

Male stylistic features included *boosters* (of course, obviously, really), *universal quantifiers* (all, always, never), *obligation modals* (must, have to), *profanity and crassness* (including sexual references), and *face threats* (commands, challenges, insults).

Average message length (in words) was also calculated for each contestant in the game play, in as much as message length has been found to correlate with gender in other forms of CMC, with males posting longer messages on average than females (Herring, 1996). Anecdotally, message length is considered a good indicator of gender identity in The Turing Game as well, although the participants interviewed by Berman and Bruckman (2001) believe that males post shorter messages than females, consistent with the stereotype that women talk more than men.

Finally, topical content was coded as female-stereotypical, male-stereotypical, or neutral for contestants' answers, as well as for the questions asked of the contestants by judges and moderators during the game play, because we observed that stereotypical questions often seemed to elicit stereotypical responses. Examples of questions and answers coded as each type are given below:

Questions and Answers

Q: How often do you buy new clothes? What was your last great "find"?
A: It was a week ago . . . I bought a beautiful black skirt. Very sexy!
   (female-stereotypical question and answer [shopping])
Q: What are your favorite porn sites on the Web?
A: Tits of the Week.
   (male-stereotypical question and answer [pornography])
Q: What's yoru fav ride at king's dominion?? [sic]
A: Hate to evade another question, but I've actually never been. I've never had a chance with school and all.
(neutral question and answer [entertainment])

Female-Stereotypical Questions

Describe a situation where you cried. (emotions)
Which type of birth control is most pleasant for you? (reproduction)

Male-Stereotypical Questions

What would be a normal setting for a spark plug gap? (cars)
What is a Flat Bastard? (computers)

Neutral Questions

What's your favorite movie? (entertainment)
If you were another creature, what creature would you be? (animals)

Statistical Analysis

Chi-squared analyses were performed to test for statistical significance of the content analysis results, and correlations between content analysis results and judges' ratings were sought. To obtain suitably conservative estimates, the Yates correction was applied in all chi-squares with one degree of freedom. In analyses where the assumptions of the chi-squared test were not met, we employed regression analysis in which the significance of individual effects was assessed by the Wald test.

Table 1
Frequencies of Nicknames

	Nickname			
Game Type	Male	Neutral	Female	All
Male games				
Real males	13	5	1	19
Real females	12	3	1	16
Female games				
Real males	1	4	11	16
Real females	2	2	18	22
All games				
Real males	14	9	12	35
Real females	14	5	19	38

RESULTS

In this section, we present the results following the order of the three research questions. The first question asks: How do contestants in gender identity games present themselves? The first means we will consider is through their choice of nicknames. Names are a conventional signal as regards gender in English-speaking cultures, and taking on a gender-specified nickname in a chat environment can be a low-cost means of "gender swapping" (Bruckman, 1993; Danet, 1998). Contestants in gender games in The Turing Game are explicitly instructed to take on a nickname consistent with the gender they are enacting, and most do. Descriptively speaking, male nicknames are preferred in male games, and female nicknames are preferred in female games. There is no significant difference in the strength of this preference according to the gender of the game, $\chi^2(1) = .04$, or according to the real-life gender of the contestants, $\chi^2(2) = 2.13$. These results are summarized in Table 1.

We also analyzed the distribution of a set of stylistic features hypothesized from previous research (for a summary, see Panyametheekul & Herring, 2003) to correlate with real-life gender. Specifically, we added together tokens of each of the male and female features listed earlier to create composite male and female scores for each contestant, for a subset of games containing two or more contestants in the corpus ($n = 14$).[9] The results, aggregated by contestant gender and gender of game, are presented in Table 2.

Although real-life females used more stylistic features when portraying online females, the overall pattern exhibited no significant differences. Neither was a preference found for the use of stylistic features appropriate to the gender of the game, $\chi^2(1) = 2.82$. Female stylistic features were displayed more frequently overall, $\chi^2(2) = 14.11$, $p < .001$, although this may be due to the fact that more female than male features were included in our coding scheme.

Table 2
Frequencies of Male and Female Stylistic Features

Game Type	Male Style	Female Style	All
Male games			
Real males	20	22	42
Real females	21	30	51
Female games			
Real males	15	35	50
Real females	24	48	72
All games			
Real males	35	57	92
Real females	45	78	123

Although we did not analyze individual features, the presence of variation across features must be noted. Absolute quantifiers (male), and modifiers, evidentials, and clausal mitigation (female) were used by male and female contestants in both male and female games. However, most instances of emoticons, laughter, thanks, and apologies were used in female games, and all but one instance of profanity and sexual language are found in male games. The following conversation from a male game involving a moderator (Anonymous) and a male contestant (Horace) illustrates the use of male stylistic features:

Anonymous:	What would you do with a million dollars?
Horace:	Get twatted. Get a flat. Go on holiday.
Anonymous:	Presuming you know the show, which actor do you like the most from Friends?
Horace:	They're all shite. Joey.

Gender style was also related to message length. Contrary to the beliefs of Turing Game participants that male messages are shorter (Berman & Bruckman, 2001), but consistent with previous findings on message length in asynchronous CMC (Herring, 1996), messages produced by real-life males are longer on average (14.8 words) than those produced by real-life females (13.0 words), especially in male games (Wald test $t = 2.58$, $p = .01$). These results are summarized in Table 3.

The last two features of game behavior we analyzed were the stereotypicality of moderators' questions and contestants' responses. Moderators in male games asked more male-stereotyped and neutral questions, whereas moderators in female games asked more female-stereotyped questions, $\chi^2 (1) = 121.4$, $p < .001$. Moreover, moderators asked more female-stereotyped questions in female games than they asked male-stereotyped questions in male games ($\chi^2 (1) = 10.14$, $p < .01$), as shown in Table 4.[10]

Table 3
Means and Standard Deviations for Message Length in Words

Game Type	*M*	*SD*
Male games		
Real males	15.8	6.7
Real females	11.8	4.6
Female games		
Real males	14.0	4.6
Real females	13.6	6.0
All games		
Real males	14.8	5.6
Real females	13.0	5.5

Table 4
Question Content Frequency and Percentages

Game Type	Male-Stereotyped	Neutral	Female-Stereotyped	All
Gender games				
Male	133 (50%)	83 (31%)	50 (19%)	266
Female	14 (8%)	44 (24%)	124 (68%)	182
All	147 (33%)	127 (28%)	174 (39%)	448

Table 5
Answer Content Frequency and Percentages

Game Type	Male-Stereotyped	Neutral	Female-Stereotyped	All
Male games				
Real males	144 (69%)	39 (19%)	26 (12%)	209
Real females	77 (69%)	24 (21%)	11 (10%)	112
Female games				
Real males	22 (27%)	11 (13%)	49 (60%)	82
Real females	27 (16%)	44 (27%)	94 (57%)	165
All games				
Real males	166 (57%)	50 (17%)	75 (26%)	291
Real females	104 (37%)	68 (25%)	105 (38%)	277

The results for the content of contestants' answers are summarized in Table 5.

Consistent with the distribution of gender-stereotyped questions, more male-stereotyped answers were given in male games, and more female-stereotyped answers were given in female games, $\chi^2(1) = 260.5$, $p < .001$. Furthermore, contestants gave more male-stereotyped answers in male games than they gave female-stereotyped answers in female games, $\chi^2(1) = 8.2$, $p < .01$. Thus, although contestants in female games were asked more gender-stereotyped questions, male games produced more gender-stereotyped responses. Moreover, real-life males gave more male-stereotyped answers than real-life females gave

Table 6
Male and Female Stylistic Features and Assessment of Gender Authenticity

	Stylistic Features		
Game Type	Male	Female	All
Male games			
Male winners	12	3	15
Female winners	12	20	32
Male losers	8	19	27
Female losers	9	10	19
Female games			
Male winners	2	9	11
Female winners	14	30	44
Male losers	13	26	39
Female losers	10	18	28
All games			
Winners	40	62	102
Losers	40	73	113

female-stereotyped answers, $\chi^2(1) = 15.5$, $p < .001$. In sum, the results for the question of how contestants behave show differences according to the gender they are performing in the game. At the same time, despite their attempts to cross gender, contestants persist in giving off cues to their actual gender through their language use.

The second research question asks what aspect(s) of contestants' self-presentation judges attend to when assessing gender authenticity. Which, if any, of the behaviors described above garner higher (or lower) ratings? "Winners" are operationalized in what follows as the contestants who received the highest score in their respective games; "losers" are all others.[11]

There is no correlation between choice of nickname and judges' ratings, $\chi^2(2) = .18$. Winners and losers of male games used male nicknames equally, and using female names in female games conferred no advantage.

There was a limited relationship between use of gendered stylistic features and ratings. Male winners avoided game-opposite gender features more than did female winners, or losers of either gender, $\chi^2(1) = 5.95$, $p = .02$. However, winners did not use significantly more gender-appropriate stylistic features than did losers overall, $\chi^2(1) = .12$. This distribution is shown in Table 6.

There was no difference in average message length between winners and losers, Wald test $t = 1.80$. However, an interaction effect occurred between ratings, message length, and real-life gender, Wald test $t = -2.56$, $p = .014$. Male winners have shorter messages than male losers, whereas the converse is true for female winners and losers. This

Table 7
Message Length in Words and Assessment of Gender Authenticity

	Message Length	
Game Type	*M*	*SD*
Male games		
Winners	12.6	5.20
Losers	15.1	6.70
Female games		
Winners	14.8	6.67
Losers	13.0	4.32
All games		
Winners	13.9	6.05
Losers	13.8	5.30

Table 8
Question Content and Assessment of Gender Authenticity in Male Games

Group	Male-Stereotyped	Neutral	Female-Stereotyped	All
Winners	92	13	5	110
Losers	89	39	25	153
All	181	52	30	263

Table 9
Answer Content and Assessment of Gender Authenticity in Male Games

Group	Male-Stereotyped	Neutral	Female-Stereotyped	All
Winners	57	7	15	79
Losers	79	34	46	159
All	136	41	61	238

is despite the fact that males in the sample have longer messages than do females overall (see Table 3). This is shown in Table 7.

With regard to stereotyped content, within the male games winners and losers were asked equal numbers of male-stereotyped questions, but winners were asked fewer female-stereotyped and neutral questions, χ^2 (2) = 19.8, p < .001. Because questioners did not know in advance who would win or lose, and because all questions were asked of all contestants in each game, this result presumably means that female-stereotyped and neutral questions were asked less often in male games with a higher concentration of winners (i.e., in games with two, rather than three or more, contestants). This distribution is shown in Table 8.

Similarly, winners of male games gave proportionately more male-stereotyped answers than did losers, $\chi^2(2) = 11.50$, p < .01. This pattern is shown in Table 9.

Table 10
Ratings of Gender Performances

Gender	Mean Rating	SD	# Performances
male → female	5.00	1.68	12
male → male	4.49	1.67	16
female → male	4.47	1.56	13
female → female	4.16	1.66	19
All males	4.71	1.67	28
All females	4.28	1.59	32
All same-sex	4.31	1.65	35
All different-sex	4.72	1.61	25
All performances	4.48	1.63	60

In contrast to the male game results, contestants' use of stereotyped answers did not affect judges' ratings in female games. No significant effect of stereotyped content was observed for winners and losers of female games, $\chi^2 (2) = 5.48$, despite the fact that contestants in female games were asked more female-stereotyped questions.

The third and final research question asks how successful contestants' self-presentation strategies and judges' assessment strategies were. To address the first part of this question, we calculated average ratings for same-sex and different-sex performances according to contestants' real-life gender. We were interested to know whether males or females were more successful at playing The Turing Game, and whether people "performing" their own gender were rated as more authentic than those performing a different gender. These results are presented in Table 10.[12]

Males and females did not receive significantly different scores for gender authenticity (Wald test, $t = -1.00$). Moreover, contrary to expectation, there was no significant difference between same-sex and different-sex performances, Wald test $t = -0.97$. The performances were rated as not very convincing overall ($M = 4.48$ on a scale of 0-10).

The judges' accuracy was calculated for games with two or more contestants by coding as "accurate" any same-sex performance rated above 5.0, and any cross-sex performance rated below 5.0. Ratings of exactly 5.0 ($n = 11$) were excluded from this calculation. For the remaining 47 gender performances, the judges' assessments of gender authenticity were incorrect 53% of the time, which is to say that they could have guessed randomly with comparable success. Given that the contestants provided linguistic cues to their actual gender, it appears that the ratings reflect not so much what the contestants actually did, as the biases and expectations of the judges.

DISCUSSION

On the basis of the above findings, we may now revisit the hypotheses set out in the beginning of the article.

> *H1a:* Contestants will choose nicknames appropriate to the gender they are performing.

This hypothesis is supported on the basis of descriptive statistics, although the finding is not surprising given that contestants were instructed to select gender-appropriate nicknames.

> *H1b:* Contestants will make use of stylistic features (hedges, boosters, apologies, profanity, etc.) appropriate to the gender they are performing.

When gender styles are considered overall, they present little evidence that people are consciously taking on the stylistic features of the gender they are performing. Although there is a nonsignificant trend for more female features to be used in female games, stylistic features—including message length—appear to map on to contestants' real-life gender more than to their performed gender, consistent with previous findings on the reliability of stylistic cues in identifying real-life gender. It may be, however, that the composite styles that we coded mask selective manipulation of individual features for strategic effect. A candidate for such manipulation might be the use of profanity and sexual crudeness, which occurs almost exclusively in male games, and which is a salient behavioral phenomenon accessible to conscious reflection.

> *H1c:* Contestants will produce topical content appropriate to the gender they are performing (that is, they will talk about topics males or females are conventionally expected to talk about, and give content-appropriate answers to questions asked by the judges and moderator).

Prompted by highly stereotypical questions, contestants of both genders give stereotypical responses that are appropriate to the gender they are performing. This appears to be the principal strategy according to which gender is consciously enacted in this chat environment.

> *H2:* Contestants will preserve some features of their real-life gender, regardless of the gender they are performing.

This hypothesis is supported for features of gender style and message length.

H3a: Contestants who choose nicknames appropriate to the gender they are performing will be rated as more authentic.

H3b: Contestants who make use of stylistic features (hedges, boosters, apologies, profanity, etc.) appropriate to the gender they are performing will be rated as more authentic.

H3c: Contestants who produce topical content appropriate to the gender they are performing will be rated as more authentic.

Neither nickname choice nor gender style correlates with ratings of gender authenticity. The significant correlation found between message length and ratings appears to be based on the popular stereotype that men talk less than women, rather than on the behavior of actual males and females. Similarly, contestants who produce stereotypical male content in male games are rated as more authentic, suggesting that judges are basing their assessments primarily on cultural stereotypes about gender. Because these are easier to fake than discourse styles, reliance on this strategy may explain the high rate of error for the judges' assessments.

H4: Contestants performing their real-life gender will be rated as more authentic than those performing a different gender.

This hypothesis is not supported. If anything, the opposite appears to be the case for the data we analyzed (see Table 10).

Further research is needed to test the possibility raised by Hills (2000), and supported anecdotally by Berman and Bruckman (2001), that contestants whose performances are highly gender stereotypical will be rated as less authentic than those whose performances are less stereotypical but still gender appropriate. In the present study, this appears not to be the case in male games, in that winners produce highly stereotypical gendered content, although our analysis did not formally distinguish different degrees of stereotypicality. However, stereotypical responses in female games may result in lower ratings, as suggested by the fact that losers in female games produce disproportionately more stereotypical content than winners. If so, this finding would be ironic, in that female game contestants are asked significantly more female-stereotypical questions—possibly as traps, on the assumption that real females would avoid the stereotypes. Both observations suggest that female behavior is more culturally stereotyped than male behavior, or at least that the judges were more consciously aware of female-gender stereotypes. A different set of expectations and standards for females and males could explain why real-life females are rated somewhat lower in their gender performances (see Table 10)—including when they are performing femaleness—than are real-life males, even when their behaviors are analogous.

CONCLUSION

In this study, we have analyzed gender performances—including deceptive performances—in a publicly available Internet chat environment designed to promote greater awareness of gender and identity. Both the performers and those who evaluated their authenticity were consciously aware of the artifice of the situation, which is characterized as a series of "games." It is possible that this conscious awareness caused them to behave and react differently from how they would in a typical chat room, where they might be less self-conscious or alert. In particular, the role of judge would appear to predispose participants toward a skeptical mindset that is presumably attenuated or lacking in casual interaction. Thus some caution should be exercised in extending the implications of these findings to other CMC contexts.

It also remains to account for why our findings differ from previous findings on what aspects of online behavior people attend to in judging gender authenticity. The studies by Savicki et al. (1999), Hills (2000), and Thomson and Murachver (2001) were carried out as experimental manipulations with conscripted subjects, whereas Turing Game participants self-selected to play the game. It is possible that the playful environment caused judges to be more receptive to behavioral stereotypes, which Herring (1998) observed to be a popular resource for role play in recreational chat environments.

The experimental studies were also based on e-mail messages, which are asynchronous, whereas The Turing Game is a synchronous (real-time) chat environment. Synchronicity—the requirement to produce and respond to messages in real time—may help to explain why stereotypes seemed persuasive to Turing Game judges, but not to the judges in previous studies. Synchronous chat imposes production and processing constraints, predisposing users toward less complex forms of expression (Ko, 1996). Gender stereotypes, which reduce the complexities of human variation to simple binaries, are easy to produce with minimal conscious attention, and require minimal attention to recognize. In other words, stereotypes, like nicknames, are low-cost conventional signals of gender identity. In contrast, discourse styles are more costly assessment signals (cf. Donath, 1999). The present findings suggest that in real-time chat in a mediated environment such as The Turing Game, linguistic cues to gender tend to be unconsciously reproduced, and largely overlooked.

These findings have implications for determining whether someone is male or female online, as well as for the phenomenological question of how successful one ultimately can be at deceiving others about gender identity through language. Although cases of successful long-term gender bending have been reported (McRae, 1996; van Gelder, 1985), a larger body of evidence points to the conclusion that conventionally gendered ways of communicating are deeply embedded in people's

social identities, and that differences tend to persist even in conscious attempts to manipulate gendered language, regardless of whether others attend to them.

NOTES

1. Participants sometimes explicitly question the gender authenticity of others in chat rooms. However, they tend to comment only on dubious performances, providing no systematic basis for comparing less successful to more successful performances.

2. "To Tell the Truth" was a classic television game show broadcast in the United States from 1956 to 1981, on which several contestants claimed to be the same person; through questioning, a panel of celebrity judges had to determine who the real person was. Contestants won prizes for "fooling" the judges. See, for example, http://www.chris-lambert.com/TRUTH/Truth.html.

3. http://www.turing.org.uk/turing/scrapbook/test.html

4. The name is further appropriate in that Turing was homosexual, and in his original proposal for the Turing Test, the human participants had to pretend to be the opposite gender (http://en.wikipedia.org/wiki/Turing_test).

5. The average game in the sample analyzed in this study involved 5.9 participants, of whom 1 was the moderator, 2 were contestants, and 2.9 were judges.

6. Only 7 gender games were played in August 1999, and 13 gender games in February 2000.

7. Games were deemed unusable if they did not provide the three types of information for any contestant, or if none of the scores had moved from the default 5.0 setting. We also excluded one game in which the only contestant self-revealed as transsexual.

8. The overall distribution of gender games played during the active period is 52% female, 46% male, and 2% other (e.g., bisexual games, or guess-my-gender without it being a specified gender game).

9. The subset comprises three games from August 1999, four games from October 1999, three games from December 1999, and four games from February 2000.

10. Table 4 does not give breakdowns for real-life gender because all questions were asked of all contestants in any given game.

11. Male games are somewhat longer, averaging 42.9 messages in contrast with 40.7 messages for female games. This may account for why more questions are asked in male games than in female games.

12. Because each game had an average of two contestants, this meant that there was generally one winner and one loser per game. This method was employed, rather than coding "winners" as those with scores above 5.0, because in some games, no contestant was scored above 5.0. The "winner" and the "loser" codes were only assigned in games with two or more contestants.

13. Only the performances of contestants in games with two or more contestants were included in these calculations. Performances that received scores of exactly 5.0 were also excluded. In addition, two outlier scores were excluded: a male-female performance that was rated .25, and a female-male performance that was rated 10.0, both by a single rater.

14. We leave aside here the complex question of whether people claiming to be their real-life gender are being themselves, as opposed to performing a different identity with the same gender, other than to note that one of the male interviewees in Berman and Bruckman (2001) claimed to perform an "idealized" version of himself, rather than his actual self—a strategy he believed resulted in higher ratings.

REFERENCES

Bell, V., & de la Rue, D. (1995). *Gender harassment on the Internet*. Retrieved May 22, 2004, from http://www.gsu.edu/lawppw/lawand.papers/harass.html

Berman, J., & Bruckman, A. (2001). The Turing Game: Exploring identity in an online environment. *Convergence*, 7(3), 83-102.

Birdwhistell, R. L. (1964, December). *The tertiary sexual characteristics of man: A fundamental in human communication*. Paper presented to the American Association for the Advancement of Science, Montreal, Canada.

Birdwhistell, R. L. (1970). Masculinity and femininity as display. *Kinesics and context: Essays on body motion* (pp. 39-46). Philadelphia: University of Pennsylvania Press.

Bruckman, A. (1993). *Gender swapping on the Internet*. Retrieved May 22, 2004, from http://www.cc.gatech.edu/elc/papers/bruckman/gender-swapping-bruckman.pdf

Butler, J. P. (1990). *Gender trouble: Feminism and the subversion of identity*. London: Routledge.

Curtis, P. (1992). Mudding: Social phenomena in text-based virtual realities. *Proceedings of DIAC92*. Retrieved May 22, 2004, from parcftp.xerox.com in pub/MOO/papers/DIAC92

Danet, B. (1998). Text as mask: Gender, play and performance on the Internet. In S. Jones (Ed.), *Cybersociety 2.0* (pp. 129-158). Thousand Oaks, CA: Sage.

Donath, J. S. (1999). Identity and deception in the virtual community. In M. Smith & P. Kollock (Eds.), *Communities in cyberspace* (pp. 25-59). London: Routledge.

Hall, K. (1996). Cyberfeminism. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 147-170). Amsterdam: John Benjamins.

Herring, S. C. (1992). *Gender and participation in computer-mediated linguistic discourse* (ERIC document ED345552). Washington, DC: ERIC Clearinghouse on Languages and Linguistics.

Herring, S. C. (1993). Gender and democracy in computer-mediated communication. *Electronic Journal of Communication*, 3(2). Retrieved May 22, 2004, from http://hanbat.chungnam.ac.kr/leejh/txt/Herring.txt

Herring, S. C. (1994). Politeness in computer culture: Why women thank and men flame. *Cultural performances: Proceedings of the third Berkeley women and language conference* (pp. 278-294). Berkeley, CA: Berkeley Women and Language Group.

Herring, S. C. (1995). Men's language on the Internet. *Nordlyd*, 23, 1-20.

Herring, S. C. (1996). Two variants of an electronic message schema. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 81-108). Amsterdam: John Benjamins.

Herring, S. C. (1998, September). *Virtual gender performances in Internet Relay Chat*. Talk presented at Texas A&M University.

Herring, S. C. (2003). Gender and power in online communication. In J. Holmes & M. Meyerhoff (Eds.), *The Handbook of language and gender* (pp. 202-228). Oxford: Blackwell.

Hills, M. (2000). *You are what you type: Language and gender deception on the Internet*. Honors thesis, University of Otago. Retrieved May 22, 2004, from http://www.netsafe.org.nz/ie/downloads/Internet_language_dissertation.pdf

Kendall, L. (1998). Meaning and identity in "cyberspace": The performance of gender, class, and race online. *Symbolic Interaction*, 21, 129-153.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, 39, 1123-1134.

Ko, K-K. (1996). Structural characteristics of computer-mediated language: A comparative analysis of interchange discourse. *Electronic Journal of Communication*, 6(3). Retrieved May 22, 2004, from http://www.cios.org/getfile\KoV6N396

Lakoff, R. T. (1975). *Language and woman's place*. New York: Harper & Row.

McRae, S. (1996). Coming apart at the seams: Sex, text and the virtual body. In L. Cherny & E. Weise (Eds.), *Wired_women* (pp. 242-263). Seattle, WA: Seal Press.

Panyametheekul, S., & Herring, S. C. (2003). Gender and turn allocation in a Thai chat room. *Journal of Computer-Mediated Communication*, *9*(1). Retrieved May 22, 2004, from http://www.ascusc.org/jcmc/vol9/issue1/panya_herring.html

Roberts, L. D., & Parks, M. R. (1999). The social geography of gender-switching in virtual environments on the Internet. *Information, Communication, and Society*, *2*, 521-540.

Savicki, V., Kelley, M., & Oesterreich, E. (1999). Judgments of gender in computer-mediated communication. *Computers in Human Behavior*, *15*, 185-194.

Stone, A. R. (1996). *The war of desire and technology at the close of the mechanical age*. Cambridge, MA: MIT press.

Thomson, R., & Murachver, T. (2001). Predicting gender from electronic discourse. *British Journal of Social Psychology*, *40*, 193-208.

Van Gelder, L. (1985, October 4). The strange case of the electronic lover: A real-life story of deception, seduction, and technology. *Ms.*, *14*, 94, 99, 101-104, 117, 123, 124.

Walther, J. (1996). Computer-mediated communication: Impersonal, interpersonal and hyperpersonal interaction. *Communication Research*, *23*, 3-43.

Susan C. Herring (Ph.D., University of California, Berkeley) is a professor of information science and linguistics at Indiana University. Her interests include computer-mediated discourse analysis, applied to gender styles, (im)politeness, (in)formality, (in)coherence, and change over time in CMC. She is Editor of the Journal of Computer-Mediated Communication*.*

Anna Martinson (M.L.S., Indiana University) is a doctoral candidate in information science at Indiana University specializing in gender, discourse, and information technology. Her dissertation focuses on ideological debates about feminism on the Web.