7 STUDY THREE: MULTIPARTY DISCOURSE IN THE BROWSER

Why do users feel OBA is "creepy"? As advertisers say, OBA simply tailors ads so that users receive more relevant advertising. Granted, this involves collecting data about online activity, but consumers willingly give data to data to first-party sites all of the time. What makes this feel different?

7.1 OBA IS "CREEPY"

Blase Ur (2012) conducted semi-structured interviews with 48 non-technical users about their attitudes and understanding of OBA. As they note, these participants are not representative of the general Internet population. The researchers were interested in collecting data that would reveal mental models of "lay people" to better understand attitudes and behaviors.

Here are a selection of participant comments from this study after they had viewed an instructional video on OBA:

"It is a little creepy... because I feel that I should get to decide what is going in and out of my computer." (Blase Ur, 2012, p. 6)

"It makes me feel very insecure. Like if this is what people can figure out about me, then what else can they get off my computer?" (Blase Ur, 2012, p. 6)

"I guess I would be more willing to do it if I had a firmer understanding of how everything worked." (Blase Ur, 2012, p. 7)

"I don't think I really noticed it... but it definitely is kind of creepy when you think about it." (Blase Ur, 2012, p. 7)

"It's kind of a creepy thought that you are being followed and monitored." (Blase Ur, 2012, p. 7)

One of the subjects,

> relating a story about how she was searching for furniture the previous night and was confused when

> her advertisements started to feature those items.... 'It's scary. It makes me nervous. I was thinking about it last night when I was searching for stuff. Like I thought how do they know all this, how do they keep track of this, how do they do this?' (Blase Ur, 2012, p. 7)

"It makes me want to go home and delete all my cookies, but then I know that's not gonna help much. It makes me mad." (Blase Ur, 2012, p. 7)

The lay person appears to have a strong emotional response to OBA. Can a theory of social communication lend insight into why?

7.2 RELATED PRIVACY RESEARCH

What makes people willing to share sensitive information online? Behavioral economists Acquisti, John, and Loewenstein (2012) find that disclosure operates under certain principles espoused by Kahneman and Tversky (1979) and Tversky and Kahneman (1981).

In an analysis of decision under risk, Tversky and Kahneman (1981) find that people people make choices by weighing their decisions in light of choices and perceived outcomes of competing choices. Choices are comparative:

They also find that choices depend on a starting reference point:

> An essential feature of the present theory is that the carriers of value are changes in wealth or welfare, rather than final states. This assumption is compatible with basic principles of perception and judgment. Our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes. When we respond to attributes such as brightness, loudness, or temperature, the past and present context of experience defines an adaptation level, or reference point, and stimuli are perceived in relation to this reference point. (Kahneman & Tversky, 1979, p. 277)

Acquisti et al. (2012) focus on the notion of a changing reference point in order to see what sorts of signals might affect a respondents willingness to disclose sensitive information. They do so by measuring people's propensity to admit to engaging in sensitive (embarrassing, unethical, and illegal behavior) in a series of surveys.

They find evidence that people will admit to having engaged in sensitive behaviors when they are lead to believe that others have admitted in similar behaviors (*herding effect*). They also examine the effect of question order on disclosure (*ordering effect*). And indeed it does. Participants in the increasing conditional were less likely to admit behaviors than in the condition where questions decreased started as very sensitive and then decreased in sensitivity. Both experiments support Kahneman and Tversky (1979) and also give insight into how and why people disclose information online.

In a study related to Acquisti et al. (2012), John, Acquisti, and Loewenstein (2009) manipulated the saliency of privacy to see whether this would affect propensity to disclose. When saliency of privacy was high, willingness to admit to engaging in sensitive behaviors decreased. And when subjects were distracted from privacy concerns, their propensity to disclose was increased.

Timing also has an effect on behavior. Egelman, Tsai, Cranor, and Acquisti (2009) studied the effect of the presentation of privacy information in the context of purchase decisions. By controlling the placement, timing, and privacy level on several consumer websites during a purchase task, participants reacted differently for the purchase of privacy-sensitive items than for items with minimal privacy concerns. In four conditions, different privacy icons were used to study both when and how icon presence affects behavior. Egelman et al. (2009) found that 1) the presence of privacy indicators influenced purchase decisions, and 2) online shoppers who were less privacy-aware, paid significantly more for privacy when privacy indicators were presented *before* visiting websites than *after* they arrived at the website.

Clearly, people are willingly disclose sensitive information online every day. And they are also sensitive to the presence of visual indicators that indicate degree of privacy. In the next section, I show that behavior is also affected by knowledge of the presence of external participants in interaction.

7.3 RELATED LINGUISTIC RESESARCH

7.3.1 *The Stateful Web*

When Tim Berners-Lee first implemented the Hypertext Transfer Protocol (HTTP) in 1990, it was to facilitate communication between a client and a server on the Internet (Cern, n.d.). HTTP functioned as a protocol to facilitate communication much as the telephone: it provided the means for a client to request information from a server via a small set of verbs such as "GET" or "POST". As a protocol, HTTP is stateless and does not require the server to track state between connections. This means, if a user requests a webpage that includes a number of embedded assets (e.g., images), the web server does not have to know anything about these assets nor does it have to track the status of those requests. In response to a request, a server generally replies with a status such as OK plus information about what sort of content is to be transferred followed by that content. In fact, an HTTP and response look something like this:

```
GET /
Accept-Encoding: gzip
User-Agent: Rested 2.3 (Macintosh; Mac OS X 10.8.0; en_US)
---blank line---
```

And a response looks something like this:

```
HTTP/1.1 200 OK
Server: gws
Set-Cookie: PREF=ID=00a5929d47bf3487:FF=0:TM=1350849438:LM
   =1350849438:S=YYHqxKmLxevdwbvo; expires=Tue, 21-Oct-2014
    19:57:18 GMT; path=/; domain=.google.com, NID=65=hS0s8
   SoRDwlN1gqEDi3Mx_rjCNUBddfp2ouOmMn4OGQBBV9
   VEejtxOuKKaIDFt5TMyrBs0ZGBZ3BH-449m2GtPjqxVsYXjuN96
   DgxaOybcNHfOjXwe2R9t6G05z3Hqhd; expires=Mon, 22-Apr-2013
    19:57:18 GMT; path=/; domain=.google.com; HttpOnly
Content-Type: text/html; charset=ISO-8859-1
Transfer-Encoding: Identity
P3P: CP="This is not a P3P policy! See http://www.google.com
   /support/accounts/bin/answer.py?hl=en&answer=151657 for
   more info."
Date: Sun, 21 Oct 2012 19:57:18 GMT
X-Frame-Options: SAMEORIGIN
X-XSS-Protection: 1; mode=block
Cache-Control: private, max-age=0
```

```
Expires: -1
---blank line---
---HTML entity here---
```

In order to accommodate stateful applications — applications that remember information about previous transactions, a small piece of data called a cookie can be attached to an HTTP request and deposited in the client web browser. This concept, envisioned by Lou Montulli of Netscape in 1994, was invented to facilitate a simple sort of memory (Schwartz, 2001). In so doing, it also enabled the sense of dialogue between server and client. Originally, cookies were intended to support basic dialogue between user and website publisher for simple transactions such as those needed to support online shopping (e.g., a cookie can be used to keep track of a user's session to include items in a shopping cart).

Though the purpose of cookies was to support the need for state between HTTP transactions, this very simple concept fostered economic and social revolution on the Web: it fundamentally changed interaction on the web from *private* to *public*. Content providers found that sites could be supported by advertising revenue simply because an advertiser with an advertising server could use cookies to decide what ad to return in response to an HTTP request. Thus, in any web page transaction, tailored advertiser content could be presented alongside publisher content.

7.3.2 *A Model of MultiParty Interaction*

The client-server interaction described above is a sort of computer-mediated communication between humans. HTTP as a machine-based communications protocol enables two parties to exchange messages. Because messages are typically simple HTTP requests with textual or other media content as a response, there is no sense of shared social context involved between a user and web provider and such dialogues are not been of linguistic interest. On the other hand, linguistic analyses of machine mediated human conversation that occur in chat rooms, texting, weblogs, etc. are a vibrant area of study in peer-review venues such as language@internet[1].

Though traditionally, the term discourse has been associated with written and spoken communication, early sociolin-

[1] http://www.languageatinternet.org/

guists such as Hymes (1974), Gumperz (1982), Goffman (1981), Garfinkel (1967) and others have extended our understanding of discourse causing us to re-consider boundaries between linguistic and non-linguistic phenomenon in social discourse.

As such, the main objects of study in sociolinguist approaches to discourse analysis are *communicative signs and their patterning* (Gumperz, 1982). Dell Hymes argues that it is not language that is fundamental in discourse, but society.

> One must take as context a community, or network of persons, investigating its communicative activities as a whole, so that any use of channel and code takes its place as part of the resources upon which the members draw. (Hymes, 1974, p. 4)

Furthermore,

> It is [...] not linguistics, but ethnography, not language, but communication, which must provide the frame of reference within which the place of language and culture is to be assessed. (Hymes, 1974, p. 4)

When considering Internet communication, it matters little whether particular interactions are "conversational" — whether responding on a forum, or requesting an article from the New York Times, or "liking" a photo, or requesting an arbitrary web page. What matters is the context; what matters is that **web interaction supports a paradigm of multi-party communication**.

The web is now a very public medium in which interactions between particular persons — or between a user and publisher — occur in a virtual space occupied by many other simultaneous participants. Though web interaction may feel very personal while sitting at ones desk at home, it is as public as conversing in a cocktail party. Moreover, interaction with a particular web site can feel very much as a private dialogue where one meets again and again to continue interacting as if no absence has occurred between sessions.[2]

[2] The hidden nature of personalization is a fascinating, related area of interest. In 2011, Eli Pariser wrote a book on the topic entitled *The Filter Bubble: What the Internet is Hiding from You* in which he describes how data mining algorithms are largely responsible for a very personalized experience on the web that influences everything from search results to whose activity is seen on Facebook.

7.3.3 *Modeling the Hearer*

Though conversation has more generally been described in dyadic (two-party) models, several notable sociolinguists have touched upon multi-party conversation and distinguished between hearer / listener roles (Bell, 1984; Clark, 1996; Clark & Carlson, 1982; Clark & Schaefer, 1987, 1992; Goffman, 1981; Hymes, 1974; Levinson, 1988; Schober & Clark, 1989). Dynel (2010) has attempted to reconcile these various classifications into a unified hearer model. In part, the need to do so stems from the analyses of new forms of discourse, such as radio, television dramas, etc.

Goffman (1981) first divided hearers into two categories: ratified and unratified.[3] Ratified hearers are those "entitled to listen to the speaker" whereas unratified are non-official participants, such as bystanders, who are present but not addressed. Ratified participants include the speaker, addressee and other "official hearers" (e.g., "third parties")[4] who are expected to follow the conversation but who are not addressed. Bell (1984) and Clark and Schaefer (1992) further distinguish unratified hearers as "bystanders" and "eavesdroppers". The primary distinction between "bystander" and "eavesdropper" is that the *speaker is aware of the presence of bystander and capacity to overhear*. (Clark & Schaefer, 1992; Dynel, 2010). See Figure 1 for a graphical representation of Dynel's (2011) synthesized model.

The distinction between bystanders and eavesdroppers serves an important distinction with respect to how speakers communicate. When the speaker is aware of a bystander, she can adjust her utterance in a number of ways. In dealing with overhearers, speakers estimate how much the overhearer can infer; then design their utterances, accordingly (Clark & Schaefer, 1992).

With respect to a shared common ground, there are two sorts of information that the speaker must consider. *Open information* is information that overhearer believes or could readily guess to be in the common ground. *Closed information* is information that the overhearer doesn't believe, and

[3] Levinson (1983), following Fillmore (1971), refers to authorized speakers and authorized recipients.

[4] Note overlapping terminology with "third party" used in a legal sense. Superficially, these appear to have the same meaning. However, behaviorally, third parties in online tracking act may across a range of hearer / listener roles.

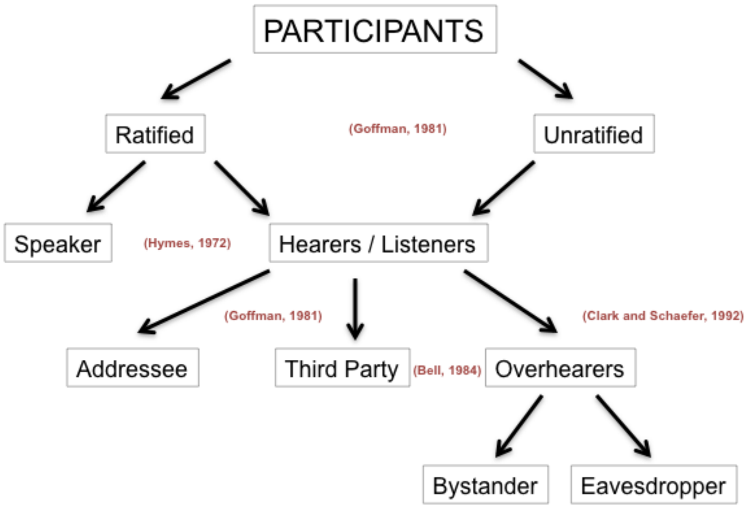


Figure 1: Modeling Hearers in Discourse (Adapted from Dynel, 2010)

could not readily guess to be in the common ground (Clark & Schaefer, 1992). It is the discrepancy between these two sorts of information that may be exploited by the speaker to affect what an overhearer (bystander or eavesdropper) might conjecture (Clark & Schaefer, 1992).

There are four main attitudes that a speaker can take toward overhearers: **indifference** (the speaker doesn't care whether the overhearer understands what she is saying or not), **disclosure** (the speaker tries to provide enough information that the overhearer can infer the right meaning), **concealment** (the speaker tries to deprive the overhearer of enough information to correctly infer what is meant — e.g., you-know-who did you-know-what to whom), and **disguisement** (the speaker attempts to conceal her meaning from the observer while also deliberately mis-representing that meaning; (Clark & Schaefer, 1992).

Thus, **audience design** can be divided roughly into addressee design, third party design, and overhearer design (Clark & Carlson, 1982). It's worth noting that participant roles are constantly negotiated between interlocutors according to regular (e.g., turn-taking) procedures. And as Dynel

(2010) notes, one role may be performed by a number of individuals simultaneously.

Online tracking represents a sort of discourse where the primary participants are a user interacting with a particular web site — which has a participant role represented by a company or organization. Other hearers are present. Perhaps, ratified and perhaps not. When overhearers are not ratified, their understanding is not grounded. What they understand is conjecture. Sometimes what they may understand may be the same as a ratified hearer, and sometimes not. In an experiment conducted by Schober and Clark (1989), ratified hearers were able to match understanding with the speaker with a very high degree of success. But despite the speaker and hearer deliberately obfuscating their speech in the presence of an overhearer, overhearers were still successful some of the time. More importantly, overhearers were also sometimes wrong.

7.4 AIMS

In this pilot, I consider whether experiment participants view advertisers as bystanders or eavesdroppers. To do this, I compare behavior across two conditions: subjects in the control condition answer a series of sensitive questions. In the treatment conditions, subjects perform the same task but are presented with visible evidence of advertisers as bystanders. This follows the general experiment design of (Acquisti et al., 2012) but poses a new dependent variable for investigation: *visual presence*.

7.5 EXPERIMENTAL DESIGN

This study has one experimental hypothesis:

Hypothesis 3
: People answering sensitive questions will behave differently in the presence of unratified participants than when simply notified in advance about the possibility of unratified participants.

In this **between** group, **single-factor**, **repeated measures** posttest-only design, the control group is asked to respond to a survey containing questions relating to sensitive activities (using questions drawn from Acquisti et al. (2012)). Participants are notified in a privacy statement that there may be ad trackers on the site. The treatment group receives exactly the same notification and survey. However, tracker presence is indicated in a visual display throughout the participant's session.

There is one **independent variable**: "visual presence". The **dependent variable** is the propensity to respond in affirmative to engaging in specific behaviors. This represents the willingness to divulge sensitive information under circumstances which may appear to be more or less private.

Following presentation of stimuli, survey data is collected for both experimental designs.

7.6 METHOD

This experimental design follow the general procedure outlined in Chapter Four.

7.6.1 *Settings and Participants*

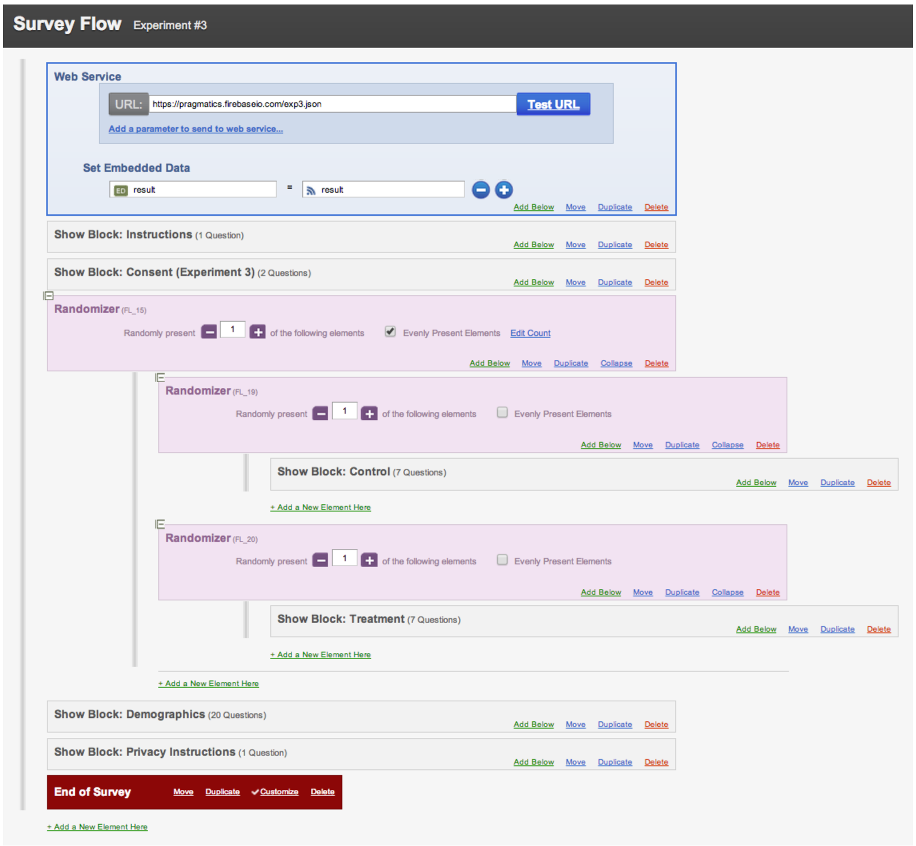
Using Amazon's Mechanical Turk, 213 workers were paid $0.15 to participate in one of two conditions of the single-factor design previously described. Participants were assigned randomly via Qualtrics block randomizer. Results were collected during the period of 7–14 October 2014. The task was expected to take approximately three minutes, though workers were allowed up to 10 minutes to complete.

7.6.2 *Procedure and Materials*

The general procedure follows that described in Chapter Four. Figure 2 is a more detailed graphical depiction of flow.

Following presentation of instructions and consent form, participants were each presented the following six question in the order given.

1. Have you bounced a check?
2. Have you cheated on a tax return?
3. Have you made a false or even somewhat inflated insurance claim?



Survey Flow Experiment #3

- Web Service
  - URL: https://pragmatics.firebaseio.com/exp3.json
  - Add a parameter to send to web service...
  - Set Embedded Data: result = result
- Show Block: Instructions (1 Question)
- Show Block: Consent (Experiment 3) (2 Questions)
- Randomizer (FL_15) Randomly present 1 of the following elements ☑ Evenly Present Elements Edit Count
  - Randomizer (FL_19) Randomly present 1 of the following elements □ Evenly Present Elements
    - Show Block: Control (7 Questions)
    - + Add a New Element Here
  - Randomizer (FL_20) Randomly present 1 of the following elements □ Evenly Present Elements
    - Show Block: Treatment (7 Questions)
    - + Add a New Element Here
  - + Add a New Element Here
- Show Block: Demographics (20 Questions)
- Show Block: Privacy Instructions (1 Question)
- End of Survey
- + Add a New Element Here

Figure 2: Experiment 3 Flow

4. While an adult, have you had sexual desires for a minor?
5. Have you had sex with the current husband, wife, or partner of a friend?
6. Have you fantasized about having violent, non-consensual sex with someone?

Possible answers were:

1. Never
2. Once or twice
3. Sometimes
4. Frequently

As in Acquisti et al. (2012), propensity to disclose is indicated by a positive answer. All positive answers are combined into a single score. Non-answers (skipped questions) count as non-disclosure. Thus, the dependent variable is binary.

7.6.3 *Instrumentation*

In addition to features provided by AMT and Qualtrics, my scripts performed the following:

1. Disable preview using a custom CSS class blocking Qualtrics controls
2. Check worker hash against a FireBase web service and request worker to return HIT if the hash is in the exclusion list
3. Add worker hash to FireBase
4. Submit results to AMT on completion of the Qualtrics survey

No other instrumentation was required for this experimental design.

7.7 DATA COLLECTED

Using a G* Power 3 f-test for linear multiple regression with a fixed model, I estimated a sample size of approximately 89 was necessary in order to detect a medium effect with power of 0.95. All surveys initiated were completed: there were no known drop-outs.

Once all assignments had been completed, (as indicated on my AMT requester dashboard), I downloaded data as a single CSV (comma separated values) file from the Qualtrics website. Data was organized such that each participant's data is on its own row.

7.8 RESULTS

There is some difference in the demographic profile between the Experiment 3 sample and the Acquisti et al. (2012) sample. That study reported a mean age of 40, 45% male, and 82% caucasian. This study was 63% male, 82% caucasian, 68% under 35, and, likely, with lower income levels (given NYT has a subscription fee for regular readers): 53% reportedly make below 30K. Regardless, comparing control group responses from Acquisti et al. (2012) with control group responses here, propensity to disclose was very similar.

Interestingly, the percentage of male respondents for Experiment 3 was higher than the aggregate population studied across all experiments in this dissertation (63% versus 53%; see **??**). This suggests that some population of workers, including females, elected not to participate in this study. I will come back to this point in a bit.

Table 1 below presents averages across all six questions. A simple chi-square test reveals no difference between groups (p = 1).[5]

There are two factors that may account for observed differences between this study and Acquisti et al. (2012). First, the population sampled here is surprisingly sensitive to privacy issues. Moreover, AMT as a platform is not seen as anonymous — as the Acquisti et al. (2012) survey may have been perceived. The mere fact that I presented an IRB form with names and contact information affected discourse context.

[5] This is with and without a bonferroni adjustment for repeated questions (p = .046)

Table 1: Raw Results (Aggregate Mean)

	Control	Treatment
Never	41.33	42.17
Once or twice	6.83	7.00
Sometimes	2.33	1.33
Frequently	.50	.50

Second, answering specific questions in the Acquisti et al. (2012) survey was completely voluntary; 'no answer' accounted for between roughly 10–15% for each question asked. This was calculated in the metric for propensity to disclose.

Selectional differences in this study are reflected in a couple of ways.

- **Population Bias.** There is no way to know if the people who elected not to participate in Experiment 3 did so because they did not want to see "sensitive" questions or because they were concerned with privacy. Regardless, there is a population bias — as reflected by the relative gender disparity between Experiment 3 and aggregate demographics across all five studies.

- **Obligation.** No one in this experiment tried *not* to answer a question. This may relate to the nature of the obligation workers feel to requesters as paid subjects.

Over the course of the past two years, tremendous change has occurred on the Internet. Savvy users, like those found on AMT, are well aware that what they say and do may be monitored. One of my participants noted in comment, "nothing is private online now due to the NSA." Events have conspired to raise the saliency of privacy issues on the Internet for even casual Internet users. Despite this, my survey reveals perceptual differences for feelings of privacy on the Internet whether connected at home or in a public setting. Figure 3 represents answers to the following two questions:

1. How private do you feel on the Internet when you are using your computer (laptop, or tablet) in public. Public means something like a coffee shop, library, school, etc.

2. How private do you feel on the Internet when you are using your computer (laptop, or tablet) at home?

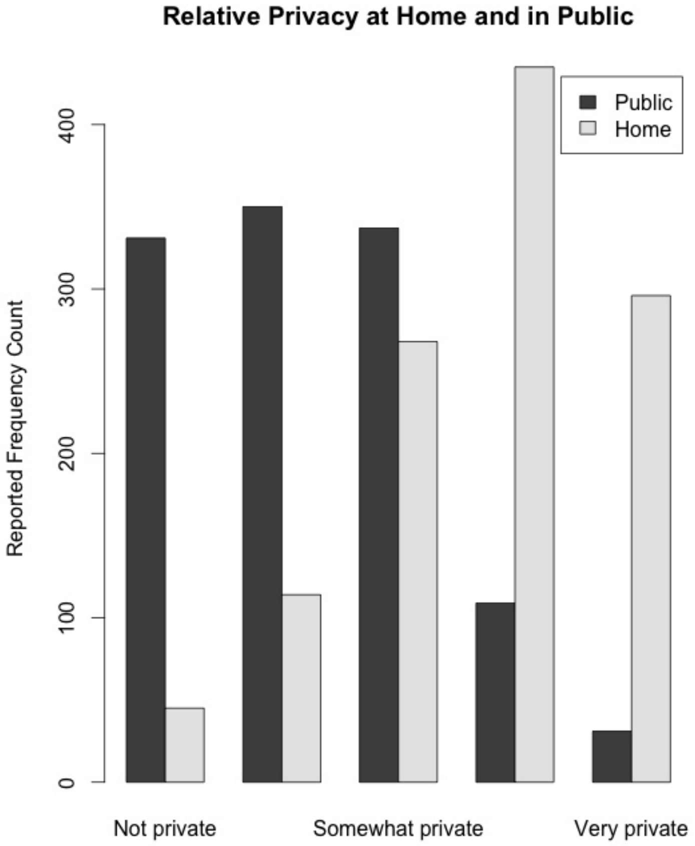


Figure 3: Feelings of Privacy on the Internet in Public and at Home

It's clear that physical setting, at least, lends some expectation of privacy whether one is on the Internet on a personal machine at home or in public. Though visual saliency was not a factor in propensity to disclose in this experiment, users have an expectation of privacy when they are not in a public setting. Given the Internet is now "public", there exists some mis-match of expectation that is not accounted for in Internet browser design.

BIBLIOGRAPHY

Acquisti, A., John, L. K., & Loewenstein, G. (2012). The impact of relative standards on the propensity to disclose. *Journal of Marketing Research*, *49*(2), 160–174.

Bell, A. (1984). Language style as audience design. *Language in Society*, 145–204.

Blase Ur, P. G. L. L. F. C. R. S. Y. W. (2012). *Smart, useful, scary, creepy: perceptions of online behavioral advertising* (Tech. Rep. No. CMU-CyLab-12-007).

Cern. (n.d.). The website of the world's first-ever web server. Retrieved from http://info.cern.ch

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech Acts. *Language*, *58*, 332–373.

Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language*, *26*(2), 209–225.

Clark, H. H., & Schaefer, E. F. (1992). Dealing with overhearers. In *Arenas of Language Use.* University of Chicago Press, Chicago.

Dynel, M. (2010). Not hearing things - Hearer/listener categories in polylogues. *MediAzioni*, *9*.

Dynel, M. (2011). On "Revolutionary Road": A proposal for extending the Gricean model of communication to cover multiple hearers. *Lodz Papers in Pragmatics*, *6*(2), 283–304.

Egelman, S., Tsai, J., Cranor, L. F., & Acquisti, A. (2009, April 4 - 9). Timing is everything? The effects of timing and placement of online privacy indicators. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009* (pp. 319–328).

Fillmore, C. (1971). Santa Cruz lectures on deixis. In *Reproduced by the Indiana University Linguistics Club 1975* (pp. 1–89).

Garfinkel, H. (1967). Studies in Ethnomethodology. *Blackwell Publishing Ltd.*.

Goffman, E. (1981). *Forms of Talk*. Philadelphia. University of Pennsylvania Press.

Gumperz, J. J. (1982). *Discourse Strategies*. Cambridge University Press.

Hymes, D. (1974). *Foundations in Sociolinguistics: An Ethnographic Approach*. University of Pennsylvania Press.

John, L. K., Acquisti, A., & Loewenstein, G. (2009). The Best of Strangers: Context-dependent willingness to divulge personal information. Retrieved from http://ssrn.com/abstract=1430482

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263–291.

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press.

Levinson, S. C. (1988). Putting linguistics on a proper footing: Explorations in Goffman's concepts of participation. In P. Drew & A. Wootton (Eds.), *Goffman: Exploring the Interaction Order* (pp. 161–227). Oxford: Polity Press.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.

Schwartz, J. (2001, September 4). Giving web a memory cost its users privacy. *New York Times*. Retrieved from http://www.nytimes.com/2001/09/04/business/giving-web-a-memory-cost-its-users-privacy.html

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.