Interaction of Email Disclosure With Invasive Survey Questions

Joe Izenman & Danish Iqbal

August 22, 2017

Abstract

Data collection and user tracking, already ubiquitous in online services, continues to grow in volume and breadth. Many contexts rely on the willingness of the user to disclose some form of uniquely identifying information. Many industries, well stocked in details about online behavior, seek to better understand their users and customers by collecting increasingly personal information about their offline life as well. It is increasingly important for both corporations and private citizens to understand the consequences of connecting offline behavior patterns to online identifiers. Our work seeks to reveal a portion of the contextual interaction between revelation of sensitive behavioral information, and the disclosure of personal contact information, in the context of a survey about socially desirable and undesirable activities.

1   Introduction

While much has been written on privacy and how individuals disclose personal detail under different circumstances, few experiments convey the type of information persons disclose when they are able to be identified.

Previous work looks specifically at the contextual environment in which personal information is disclosed. It has generally been found that in mediums signaling some degree of being unprofessional, responders are more likely to disclose embarrassing details about themselves. It is also found that signaling privacy may change how questions are responded to. Our work builds on this body of experiments to see the type of responses provided when personally identifiable information is asked for.

2   Background

While our experiment isn't conducted in the presence of an interviewer, divulging of sensitive information is dependent on context. For example, in the case of surveys administered by physical interviewers, Tourangeau and Yin (Tourangeau and Yang 2007) summarize that respondents are more likely to report sensitive information when questions are self administered, rather that being administered by an interviewer. A survey responder is less likely to disclose information if they interact with someone (such as an interviewer) who may learn something embarrassing or something that could be used to punish the survey responder.

Tourangeau and Yan also write that using familiar wording can increase reporting as well as question-wording strategies. For example, in collecting income, financial, and other numerical information, researchers sometimes use a technique called unfolding brackets to collect partial information from respondents who are unwilling or unable to provide exact amounts.

How much personal information is given depends on the context it is delivered through. Our experiment borrows some methodology from "Strangers on a Plane" (John, Acquisti, and Loewenstein 2011). John, Acquisti, et al conduct a series of experiment where the context in which sensitive questions are asked is changed. A total of four experiments are conducted. The first experiment changes how a sensitive question is asked, whether directly or indirectly. The experiment reveals that a responder is more likely to divulge private information when a question is asked indirectly. Therefore, the way a question is worded may get a surveyor more responses about personal information.

Other experiments are conducted in which responders answer personal questions on a professional-looking and an unprofessional looking website versus a baseline website. They found that responders were more likely to reveal personal information in the unprofessional-looking website than they were in the professional. Similarly, when such questions were asked on different types of survey interfaces (unprofessional vs professional), responders were more likely to have answered intrusive questions in the affirmative in the unprofessional looking survey.

In their final experiment, John, Acquisti, et al cue the responder to privacy concerns from the beginning and then see how the survey is responded to. They find that the difference in how unprofessional and professional surveys are responded is eliminated when privacy concerns are relayed in the beginning. Signaling privacy concerns prior to a survey may change how questions are responded to, a characteristics that we hope to find in our experiment.

Further research on privacy and revelations (Spiekermann, Grossklags, and Berendt 2001) finds that stated preferences for privacy may not align with actual behaviors. They conduct an experiment where users privacy preferences are compared to their interactions on shopping websites. In their experiment, users were asked to fill a questionnaire regarding their privacy preferences. Then users were given one of two privacy statement; one harsh and one soft, mentioning that their navigation patterns would be recorded during the shopping process. Participants had the option to leave at any time while shopping.

There was generally a correlation between users' navigation patterns and their privacy preferences. However, the same was not seen when it came to revealing information about themselves. User were more likely to reveal information about themselves, in contrast to their indicated privacy preferences.

Much of how privacy-related questions are answered depends on the mechanism used for asking them. Knijnenburg and Kobsa (Knijnenburg, Kobsa, and Jin 2013) write that users tend to forgo their privacy calculus when presented with filled auto-completed forms. In an experiment they present users with three different types of auto-complete forms and observe how likely users are to use the auto-complete options. Despite selective privacy preferences, users are likely to over-disclose information when using the traditional auto-complete present on many web browsers.

The level of personal information revealed may be dependent on the audience. Quian and Scott look at bloggers and how much of their private information they choose to disclose. They find that bloggers revelation of personal details is contingent on how well they know their target audience. The more personally acquainted a blogger is with their audience, they more likely they are to disclose personal details about themselves.

Finally, revelation may be dependent on how you see your traits vis-a-vis societal acceptance of norms. Huberman, et al find that responders are less likely to disclose information when their response has higher deviation from the norm. They conduct an experiment where experimenters auction personal information about themselves. They find that when a trait is more socially undesirable, the individual demands more compensation for revealing it. For example, the more a person weight was above the median, the higher the compensation they would demand. Their work shows us how different social scenarios and indicators may be valued differently by how much the trait is considered a deviation from the norm.

Our experiment touches on many of these issues as users are asked personal questions about social behaviors and are asked to provide personal contact information.

3 Experiment

3.1 Overview

The specific goal of this study is to examine the interaction–in both directions–of personal identity disclosure on the response to a variety of questions. How do revelations about one's past behavior, initially made anonymously, affect their willingness to voluntarily identify and permit future contact? And, conversely, how do the responses to the same kinds of questions change when the respondent has already provided that same information?

The core of the design is to ask respondents for the answers to a variety of questions, which have been pre-classified by their sensitivity and perceived invasiveness, with the intent to answer the following high level research questions:

- RQ1: What is the impact of the timing of our identity disclosure request on all responses?
- RQ2: What is the impact of question types on question responses?
- RQ3: What is the impact of question responses on later identity disclosure?
- RQ4: What is the impact of prior identity disclosure on question responses?

The answers to these questions, taken together, can help compose a useful picture of the fundamental interactions between behavioral honesty and anonymity.

3.2 Question Selection

The basis of our analysis is a framework of questions determined to be at different levels of "sensitivity". Tourangeau and Yan (2007) offer several models of question sensitivity, including intrusiveness and threat of disclosure. Here we focus on their third notion, that of "social desirability." Given clear social norms about specific behavior, conforming to those norms is socially desirable, and diverging from them is undesirable. Questions that inquire directly about socially undesirable behavior are perceived as more sensitive. In particular, this perception of sensitivity is linked to the an affirmative admission. Someone who has, in truth, not violated a social norm is unlikely to find the question sensitive—that is, unless they are offended by the nature of the question itself, violating intrusiveness rather than desirability.

There is some concern that a clear disparity in the desirability of behaviors mentioned in the survey may encourage response bias. If some questions are about blatantly positive behaviors, while others address extreme violations of social norms, respondents may be more likely to cater their answers to perception of these extremes. To combat this, we narrow the gap in social desirability, by focusing on desirable and undesirable forms of a single kind of behavior. In this case, all questions formulated are on different kinds of lies and deception. Social norms consider lies to be generally undesirable, but so-called "white lies", are much more accepted.

Focusing on a narrow band of behavior also aids in obfuscating the intent of the survey. We are able to introduce the survey with a plausible research topic on general deceptive behavior in the population rather than a random assortment of behaviors that leaves a subject curious, and searching for the pattern.

3.2.1 Pre-Survey

Central to a successful study of these issues is the classification of behaviors as either desirable or undesirable. To develop the question list for our main experiment, we conducted a pre-survey, establishing a statistical basis for our social desirability rankings, and helping avoid inserting our own biases and social constructs into the classification process.

3.2.1.1 Methodology

An initial list of twenty different kinds of lies was composed by the researchers (enumerated, with results, in Table ??? below). Some of these behaviors were deliberately selected based on prior belief regarding their social desirability, in the hopes of guaranteeing a wide range of options. However, these priors were not incorporated into later analysis, to allow for the possibility that our own notions of social desirability may not be shared by the population at large.

In order to prevent question overload, each respondent was shown a randomly selected subset of five behaviors. Each question asked whether or not they believed the given behavior was considered socially acceptable (Fig ???). All responses were then aggregated, and each behavior analyzed using a binomial test. The null hypothesis of this test was a probability of 0.5—that each behavior would be divided evenly between desirable and undesirable. Any questions with a statistically significant divergence (alpha = 0.05) from the

middle would be flagged with the majority vote. These questions would be used in the main survey, unless a severe class imbalance between question types arose.

Lying to a sexual partner, about a sexually transmitted disease.

Socially acceptable

Not socially acceptable

Embellishing minor characteristics, while on a date.

Socially acceptable

Socially unacceptable

Figure 1: Pre-survey question sample

Since demographic details or specific subgroups are not the focus of this study, pre-survey respondents were recruited, and compensated for their time, via Amazon Mechanical Turk. In an attempt to distill somewhat consistent social norms from a single culture, subjects were restricted to English-speaking, U.S.-based workers.

3.2.1.2 Results

253 pre-survey responses were collected, with an average of 63.25 responses per behavior. Due to random assignment, some behaviors were seen more often, and some less (Table ???), but all hovered in an acceptable range.

Table 1: Presurvey Response Rate

Total Responses	1265.00
Avg. Response Rate	63.25
Min. Response Rate	51.00
Max. Response Rate	77.00

Analysis provided sixteen statistically significant behavior evaluations. Ten were deemed socially acceptable, and six unacceptable (Table ???, Figure ???). Some of these results were within our expectations—the socially acceptable nature of encouraging belief in Santa Claus is no surprise, for example.

Of particular note was the dichotomy of lying about income (acceptable), and lying on a tax return (unacceptable). From an information perspective, these are essentially the same deception. They vary only in who they are told to: a non-specific, general audience, or the government.
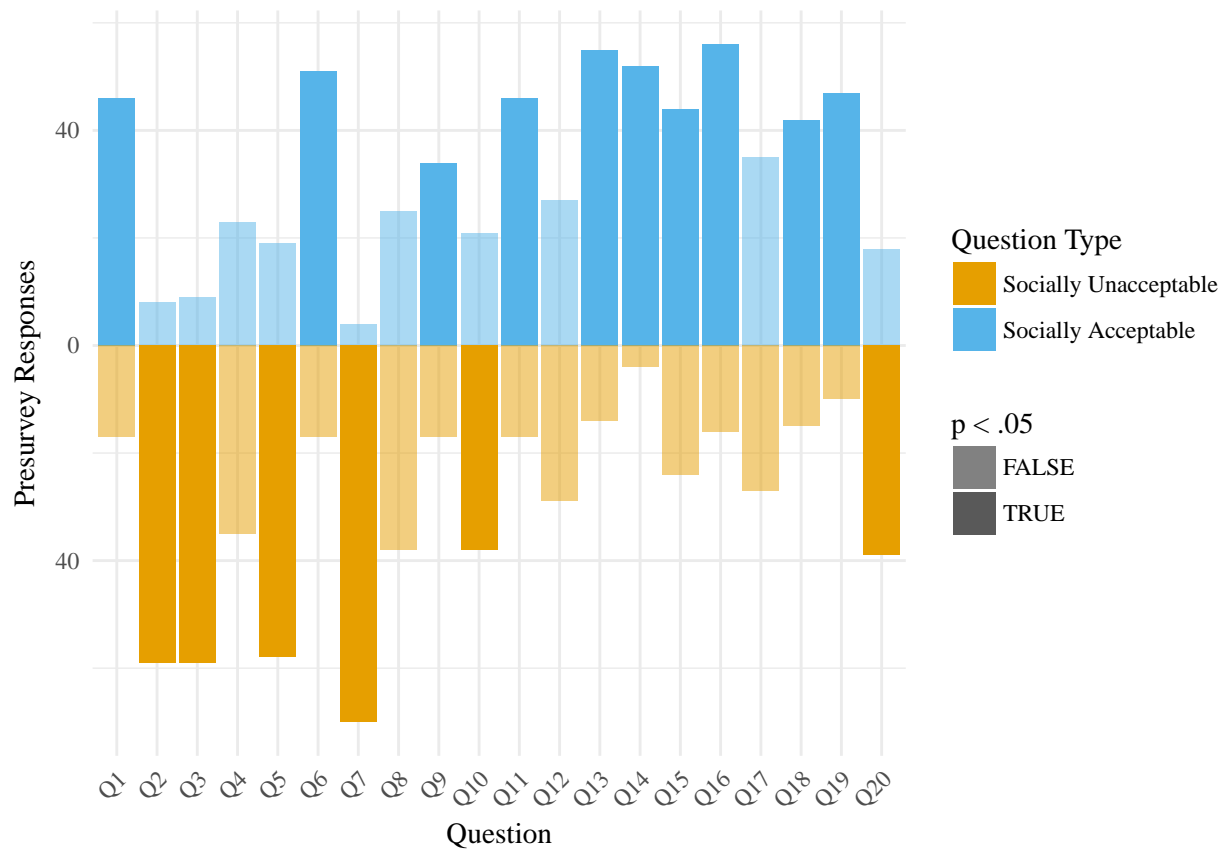
4

Figure 2: Presurvey Responses

	Question	Acceptable	Significant
1	Lying about your income.	Y	***
2	Lying on an insurance claim.	N	***
3	Lying on a tax return.	N	***
4	Lying about your age on a dating website.	N	-
5	Lying about your age to secure alcohol or other controlled product.	N	***
6	Lying about your contact information to avoid further interactions with someone.	Y	***
7	Lying to a sexual partner, about a sexually transmitted disease.	N	***
8	Lying to a law enforcement officer, to protect yourself.	N	-
9	Lying to a law enforcement officer, to protect a friend or family member.	Y	**
10	Lying to a sexual partner, about past sexual activity.	N	**
11	Lying to your friends about your sexual activity.	Y	***
12	Lying to a police officer about your speed at a traffic stop.	N	-
13	Lying to a child, to avoid a difficult conversation.	Y	***
14	Lying to a child, to maintain a tradition (such as Santa Claus or the Easter Bunny).	Y	***
15	Lying about a food allergy at a restaurant, to avoid an ingredient you dislike.	Y	**
16	Lying to a partner or friend about their appearance.	Y	***
17	Lying about an illness, to take a day off work.	Y	-
18	Lying to a friend to protect their feelings.	Y	***
19	Embellishing minor characteristics, while on a date.	Y	***
20	Lying on a resume or in a job interview.	N	***

Table 2: Presurvey Questions and Results

3.2.2 Question Presentation

There are several possible ways to form questions from a list of behaviors. John et al. (2010) compare direct to indirect questioning. Rather than directly asking about a given behavior, their method of indirect questioning provides multiple locations to rank the behavior ethically, based on whether the respondent has performed that behavior in the past. They find that this method increases admission rates for questionable behaviors.

For purposes of our study, a higher baseline admission rate is not necessarily beneficial. Among our research goals is an investigation of how prior email disclosure impacts admission rates of socially undesirable behaviors. For this reason, and for simplicity of analysis, we have elected to use direct questioning. Each behavior is presented as a simple yes/no question, of the format, "Have you, or would you, lie about x?" The inclusion of "or would you" attempts to account for cases where the respondent has not, in the course of their normal life, been presented the opportunity for the behavior. For example, someone who has never suffered from a sexually transmitted disease would not have had the opportunity to lie about one.

3.3 Main Survey

Unlike the pre-survey, we elected to present the full course of available questions to all respondents. By ensuring that everyone sees the full spread of sensitive questions—in addition to the others—we maximize the possibility of seeing a treatment effect between the groups. There is some imbalance between the groups—10 desirable behaviors, and 6 undesirable—but this was not deemed extreme enough to merit editing of the list.

The survey begins with a brief description of the purpose of the survey, and instructions on the remaining survey. It is presented as as study of "prevalence of different forms of lying and deceptive behavior in the

population." Visitors are instructed to answer yes or no, based on past behavior, and to truthfully consider any behaviors which they have not yet had the opportunity to enact.

Once the survey begins in earnest, each respondent is randomly placed into one of two groups, referred to hereafter as the "pre" and "post" groups. All are asked optionally for their email address, nominally for the purpose of conducting future follow-up research.

Have you, or would you lie to your friends about your sexual activity?

Yes

No

Have you, or would you lie about a food allergy at a restaurant, to avoid an ingredient you dislike?

Yes

No

Figure 3: Main survey question example

Are you willing to be contacted for a potential follow-up interview? If so, please provide a valid email address.

Figure 4: Email question example

As will be addressed in section 5 below, email is an imperfect proxy for personally identifiable information. Despite this, we determined that it would be sufficiently revelator to provoke possible discomfort, while also making sense as a plausible query to attach to an otherwise anonymous survey, from the respondents' perspectives.

The pre group is presented with this request for email immediately after the instructions. Then both groups are given the behavior questions, randomly ordered, and presented as the simple yes or no responses as described above. Finally, the post group is asked for their email address, using an identical request as has already been seen by the pre group.

To maximize usable data from our limited sample size, we eschew a dedicated control group. Instead, the post group serves as a control for questions about the effect of email disclosure on question response—they

have not yet been asked for identification when they are answering the questions, and thus would be have as someone who is never so asked—and the pre group serves as a control for questions about the effect of question response on email disclosure—likewise, they have not yet seen questions prior to their decision to disclose their address.

As with the pre-survey, respondents were recruited from Mechanical Turk, and paid an appropriate market-rate fee, in keeping with estimated task time, and the national minimum wage, with the hope of securing high quality results.

4 Results

4.1 Hypotheses for Analysis

Based on the original research questions, and the specific formulation of the questions as described above, we have developed a specific sequence of hypotheses to test on the acquired survey data (Table ???).

4.1.1 What is the impact of the timing of our identity disclosure request on all responses?

- **H1:** Assignment to the pre or post group will affect email disclosure rates.
- **H2:** Assignment to the pre or post group will affect the rate of "Yes" answers to all behavior questions.
- **H3:** Assignment to the pre or post group will affect the rate of "Yes" answers to questions classified as sensitive

4.1.2 What is the impact of question types on question responses?

- **H4:** The classification of a question as sensitive will impact the number of "Yes" answers it receives.

4.1.3 What is the impact of question responses on later identity disclosure?

- **H5:** In the post group, the number of "Yes" answers to all behavior questions will impact the email disclosure rate.
- **H6:** In the post group, the number of "Yes" answers to sensitive behavior questions will impact the email disclosure rate.

4.1.4 What is the impact of prior identity disclosure on question responses?

- **H7:** In the pre group, the email disclosure rate will impact the number of "Yes" answers to all questions.
- **H8:** In the pre group, the email disclosure rate will impact the number of "Yes" answers to sensitive questions.

4.2 Responses

The main survey received 301 total responses, all recruited from Mechanical Turk. 215 respondents provided email addresses in the survey. One additional respondent reached out, post-survey, to report a problem providing their email address, but it was not clear which group or set of responses they were linked to, so that disclosure is not included here.

Table 3: Main survey response rate

Total Responses	Emails Collected	Avg Yes Answers (max: 20)
301	215	7.16

The pre and post groups were created via simple random assignment, wherein each respondent's assignment was independent of all others. As such, the two treatment groups are unevenly split—46/54 rather than 50/50. The proportions are close enough to 50% that sample size is not a concern.

Table 4: Treatment summary

Treatment Time	Total Responses	Emails Collected	Avg Yes Answers (max: 20)
Post	137	94	6.95
Pre	164	121	7.33

4.3 Hypothesis Testing

All hypotheses were analyzed using linear regression models. However, since no demographic covariates were included, the regressions are all essentially t-tests.

4.3.1 Impact of disclosure request timing

The first analysis covers all respondents, based on group assignment, pre or post. The survey results show no significant difference in any responses, caused by the timing of the request for email disclosure. Disclosure rate, admission to behaviors in general, and admissions to socially undesirable behaviors on sensitive questions all show no notable effect.

Table 5: Impact of disclosure timing

	Dependent variable:		
	Email disclosure	Yes answers (all questions)	Yes answers (sensitive questions)
	(1)	(2)	(3)
Email request timing (post)	−0.052	−0.270	0.011
	(0.052)	(0.422)	(0.171)
Constant	0.738***	7.374***	1.337***
	(0.035)	(0.284)	(0.115)
Observations	301	297	298
R^2	0.003	0.001	0.00001
Adjusted R^2	−0.0001	−0.002	−0.003
Residual Std. Error	0.453 (df = 299)	3.620 (df = 295)	1.469 (df = 296)
F Statistic	0.973 (df = 1; 299)	0.408 (df = 1; 295)	0.004 (df = 1; 296)

Note: *p<0.1; **p<0.05; ***p<0.01

4.3.2 Impact of question type on responses

Our second analysis steps away from each respondent's survey as a unit, and focuses on each individual question, across all responses. Sensitivity of the question does show a highly significant ($p < 0.01$), medium sized ($R^2 = 0.127$) effect on affirmative admission rates. The negative direction indicates that respondents are less likely to admit to socially desirable behaviors. Figure 5 demonstrates the clear difference between the sensitive and non-sensitive questions.

Unfortunately, this study is not equipped to determine whether the effect is due to a reduced willingness to admit behaviors of this nature, or the population at large is genuinely less likely to participate in said behaviors. Given that the evaluation as undesirable is rooted in social norms, the latter seems likely, but any attempt at a ground truth measure of most of these behaviors would be reliant on voluntary disclosure, and thus largely unreliable.

Table 6: Impact of question type on responses

	Dependent variable:
	Yes answers
Question marked as sensitive	-0.366^{***}
	(0.014)
Constant	0.592^{***}
	(0.008)
Observations	4,812
R^2	0.127
Adjusted R^2	0.127
Residual Std. Error	0.465 (df = 4810)
F Statistic	699.166^{***} (df = 1; 4810)
Note:	*p<0.1; **p<0.05; ***p<0.01

4.3.3 Impact of question responses on email disclosure

Next, we focus on the post group, to test whether response patterns in the survey questions affect later willingness to disclose identity. Our results show no significant relationship between behavior admissions—desirable or undesirable—and eventual email disclosure. In fact, in both cases an increase in "Yes" response rate actually *increased* disclosure, though not in a statistically meaningful way.

4.3.4 Impact of email disclosure on question responses

Finally, we turn to the pre group and reverse the last analysis, instead testing whether prior disclosure of identity information affects later behavior admission response. Again, we find no significant change in the response rate to either questions at large, or sensitive questions in particular, dependent on email disclosure.

4.4 Discussion

In general, this study found no interactions of note between email disclosure and behavior admission. Admitting to socially undesirable behaviors did not affect a respondent's willingness to be provide contact information for future research. Conversely, having already provided such information did not affect their admissions.
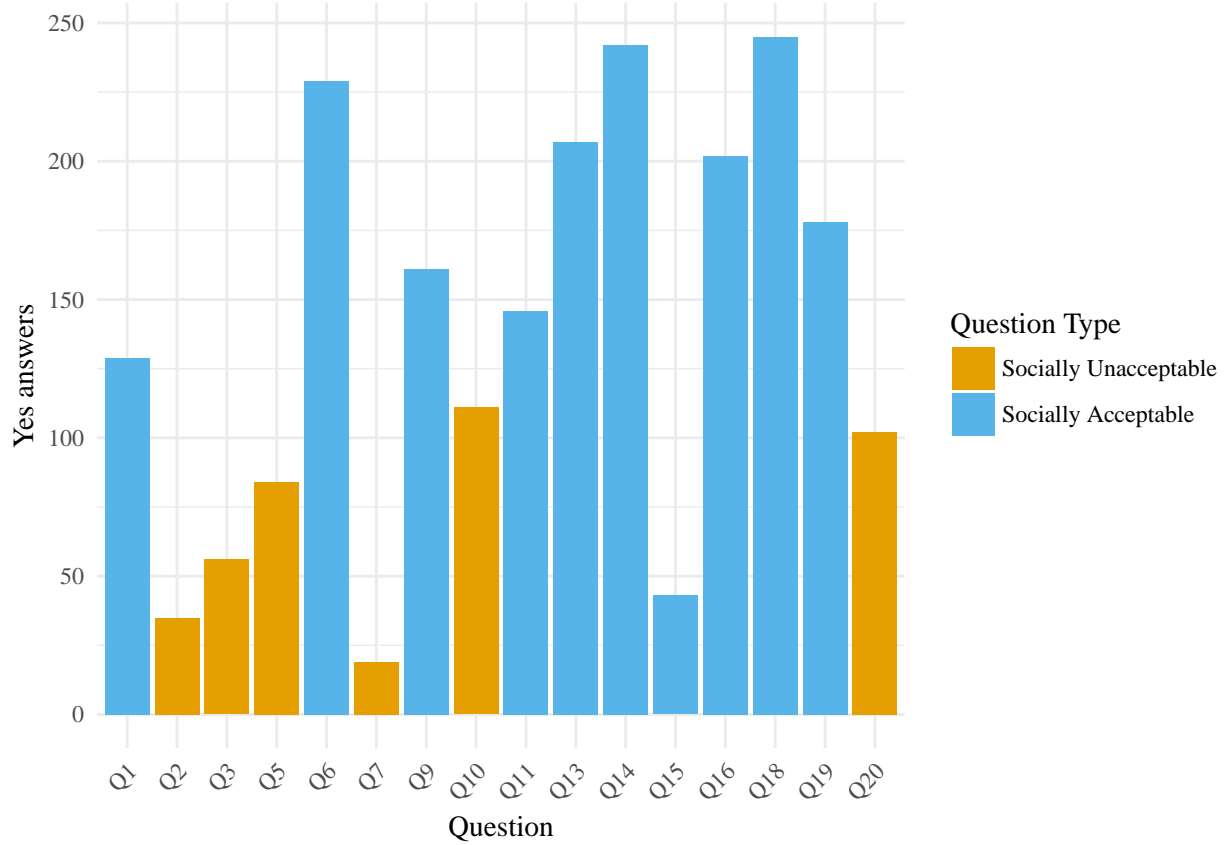
Figure 5: Question types and responses

Table 7: Impact of questions on email disclosure

	Dependent variable:	
	Email disclosure	
	(1)	(2)
Yes answers (all questions)	0.011	
	(0.011)	
Yes answers (sensitive questions)		0.015
		(0.027)
Constant	0.614***	0.668***
	(0.087)	(0.054)
Observations	134	135
R^2	0.008	0.002
Adjusted R^2	0.0005	−0.005
Residual Std. Error	0.462 (df = 132)	0.466 (df = 133)
F Statistic	1.064 (df = 1; 132)	0.330 (df = 1; 133)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 8: Impact of email disclosure on questions

	Dependent variable:	
	Yes answers (all questions)	Yes answers (negative questions)
	(1)	(2)
Email disclosure	0.215	0.070
	(0.640)	(0.260)
Constant	7.214***	1.286***
	(0.552)	(0.224)
Observations	163	163
R^2	0.001	0.0004
Adjusted R^2	−0.006	−0.006
Residual Std. Error (df = 161)	3.574	1.454
F Statistic (df = 1; 161)	0.113	0.072

Note: *p<0.1; **p<0.05; ***p<0.01

Our single statistically significant result, that question sensitivity affects response rate, is limited in its utility, because it is not clear whether the actual cause of lowered response rate is the sensitivity of the question, or the root social norm driving its evaluation as sensitive.

In section 5 we review several possible reasons why this study may have failed to find any interactions, beyond the clear option that there simply are none.

5 Problems

Though we accept the results of our study as statistically valid, and would hesitate to re-run the survey in search of better results, it remains possible that treatment effects do exist in at least some of these cases, but were not revealed by our research efforts. Several potential issues with the experimental design may have worked to suppress inherent differences between the pre and post groups.

First, the behaviors presented may be too close together. The statistical significance measured on the pre-survey does not necessarily indicate just how far apart in desirability the different questions are. Since all the behaviors are of the same general type—lying and deception—if the questions on each side of the line are too similar on a respondent's internal social desirability scale, the effect size between them would be very small, requiring a larger sample size to detect. The binomial test does calculate an estimated probability that could be used to rank behaviors, but our relatively short question list did not lend itself to this.

This effect could likely be mitigated by a more thorough pre-survey. Expert opinions may be solicited, and pre-existing research consulted, on a wider range of deceptive behavior. With a larger number of options, selections from the pre-survey results could be more discerning, filtered to only utilize those with a more extreme effect, while ignoring those questions likely to be just over the line of statistical significance.

Additionally, deceptive behavior in general has a negative connotation, and bias toward the undesirable. Since the general topic of the survey was revealed to the respondent prior to the first question, even the users in the pre group would have had a general idea that they may be asked sensitive questions, prior to being prompted for their email address. Future efforts ought to be more vague about the nature of the behaviors to be solicited in the main body of the survey, to avoid this bias.

Finally, email addresses may simply not be personal enough to prompt a reaction. This is a particular risk for a study recruiting from Mechanical Turk. Many users of such modern internet tools maintain

secondary, anonymous email addresses, not associated with their identity or professional lives, strictly for use when signing up for online services. This serves both to protect their identity and evade inbox clutter.

Review of message boards in use by "Turkers" shows several individuals who use these secondary addresses explicitly for surveys that request contact information for follow-up. Without a direct association to their identity, respondents may feel no qualms about divulging this detail, regardless of the questions asked and answered.

This suggests either that more personal information should be solicited—within reason, and within Amazon's usage policy—or that Mechanical Turk is not the appropriate recruitment venue for a study of this type. For example, a more traditional university recruitment would permit the study to request the respondent's official email address, guaranteeing identifiability in a way that is impractical with a broad, truly anonymous sample.

6 Future Work

In addition to addressing flaws in the original design, several extensions are possible to gain deeper insight into these interactions—assuming for the moment that such interactions do exist.

Treatment of social desirability as a strictly binary consideration limits the amount of detail that can be solicited. Forcing a yes/no answer can be useful when trying to maximize observable effect, but use of an ordinal scale provides room to quantify the level of desirability in each question answer and overall survey response. This provides a more nuanced view of the effect on email disclosure.

This task of assigning desirability scores to each individual survey might be accomplished in one of two ways. If the binary evaluation of each behavior remains important, each respondent could instead be shown a random subset of questions. A given survey response could then be scored based on how many of the presented questions were rated undesirable.

Additionally, as mentioned above, starting from a broader set of behaviors, and a more thorough pre-survey, would provide more insight into just how far any given behavior is from the center line of desirability, providing the researchers with an ordered ranking of behaviors.

References

John, Leslie K., Alessandro Acquisti, and George Loewenstein. 2011. "Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information." *Journal of Consumer Research* 37.

Knijnenburg, Bart P., Alfred Kobsa, and Hongxia Jin. 2013. "Counteracting the Negative Effect of Form Auto-Completion on the Privacy Calculus." In *Proceedings of the 34th International Conference on Information Systems.*

Spiekermann, Sarah, Jens Grossklags, and Bettina Berendt. 2001. "E-Privacy in 2nd Generation E-Commerce: Privacy Preferences Versus Actual Behavior." In *Poceedings of the 3rd Acm Conference on Electronic Commerce*, 38–47.

Tourangeau, Roger, and Ting Yang. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133 (5): 859–83.