

The Impact of Relative Standards on the Propensity to Disclose

Forthcoming, Journal of Marketing Research, 2011

Two sets of studies illustrate the comparative nature of disclosure behavior. The first set investigates how divulgence is affected by signals about others' readiness to divulge, and shows a "herding" effect: Survey respondents are more willing to divulge sensitive information when told that previous respondents have made sensitive disclosures (Study 1A). We provide evidence of the process underlying this effect and rule out alternative explanations, by showing that information on others' propensity to disclose affects respondents' discomfort associated with divulgence (Study 1B), but not their interpretation of the questions (Study 1C). The second set of studies investigates how divulgence is affected by the order in which inquiries of varying intrusiveness are made, and suggests that divulgence is anchored by the initial questions in a survey; people are particularly likely to divulge when questions are presented in decreasing order of intrusiveness, and less likely when questions are presented in increasing order (Study 2A). We show that the effect arises by affecting people's judgments of the intrusiveness of the inquiries (Study 2B). The effect is altered when, at the outset of the study, privacy concerns are primed (Study 2C), and when subjects are made to consider the relative intrusiveness of a different set of questions (Study 2D). This research helps to understand how consumers' propensity to disclose is affected by continual streams of requests for personal information, and by the equally unavoidable barrage of personal information about others.

Keywords: Survey Design, Self-Disclosure, Electronic Commerce, Privacy

¹ Alessandro Acquisti is an Associate Professor of Information Systems and Public Policy at the Heinz College, Carnegie Mellon University. Leslie John is a doctoral candidate in Behavioral Decision Research, Department of Social and Decision Sciences, Carnegie Mellon University. George Loewenstein is the Herbert A. Simon Professor of Economics and Psychology in the Department of Social and Decision Sciences, Carnegie Mellon University.

The central thesis of this paper is that disclosure behavior is comparative in nature: people's willingness to divulge sensitive information depends on judgments that are inherently comparative, such as signals about others' readiness to divulge, or the order in which inquiries of varying intrusiveness are made. In seven studies, we investigate disclosure behavior by asking subjects personal, and often incriminating, questions under different circumstances and monitoring their propensity to respond affirmatively – to admit that they have engaged in sensitive behaviors.

Narrowly, this research attempts to shed light on the question of when survey respondents will reveal personal information in response to intrusive, or sensitive, questionnaires. This issue has received considerable attention, both in the marketing literature (e.g., Reinmuth and Geurts 1975) and in the literature on survey methodology (e.g., Tourangeau and Ting 2007), but is also directly relevant to contemporary phenomena on the internet: Advances in information technology have been a boon to marketers, who can use personal information to tailor messages to individual consumers (Blattberg and Deighton 1991). However, information technologies deemed too intrusive by consumers can elicit reactance (White 2004), countering the marketing benefits of those tools. To predict and make sense of consumers' reaction to modern marketing strategies, we need to understand how they respond to the continual stream of requests for personal information that is an unavoidable feature of the internet, as well as to the equally unavoidable barrage of personal information disclosed about others.

Beyond this narrow focus, we believe (and have designed our experiments accordingly) that the research reported here addresses larger issues relating to consumer privacy. Although numerous privacy-related decisions do not involve responses to intrusive questions, many of the causal factors we examine are common to disparate privacy-related issues. For example, people probably share information on social media sites in part depending on what other people do (the focus of our Studies 1A-C), and in part depending on what they have themselves previously revealed (Studies 2A-D). Because examining responses to intrusive survey items is an effective and efficient methodology for testing the impact of situational factors on disclosure of information, much prior and contemporary research on privacy, including our own, has adopted such a methodology (e.g., John, Acquisti, and Loewenstein 2011).

A first set of studies (Studies 1A-C) examines the impact of receiving information about others' disclosures on one's own propensity to disclose. Online surveys often provide information on other

respondents' answers, including their propensity to answer; we study how such feedback influences subsequent responses. These studies thus provide clues about how self-disclosure might be affected by the growing availability of friends' and strangers' personal information on the internet. This effect had proved so far elusive in the literature. Ong and Weiss (2000) were surprised to find that “normalization” – suggesting that a behavior is reputedly commonplace or rare – had *no* impact on the propensity to admit to behaviors carried out in private. The authors attributed the null result to “weak implementation.” In a prior, related paper, Moon (2000) found that people reciprocated the revelation of intimacies, even from a computer. In Moon's experiment, participants were more likely to disclose when interviewed by a computer that prefaced questions with information about itself. This effect is related to our study. Whereas Moon's result could result either from reciprocity (the tendency to respond to another's revelations with one's own) or from herding (the tendency to conform to the norm set by others' behavior), our results are only interpretable as a herding effect, as we focused on the impact of *strangers'* revelations (with whom no interaction or reciprocity is likely) on our propensity to reveal sensitive information.

Moon also found that participants were more likely to disclose to a computer that had “warmed” them up with introductory questions, relative to participants who had answered the same questions, but on *another* computer. Since in Moon's study all participants actually answered the same questions in the same order across the conditions, the result evokes, but does not resolve, the question we try to address in this manuscript: how is people's propensity to disclose affected by the order in which inquiries of varying sensitivity are made? Do consumers reveal more when they are first asked unintrusive questions and are “warmed up” as the questions become more intrusive? Or does asking questions in a progressively intrusive order cue consumers to perceive subsequent questions as sensitive, causing them to “clam up,” relative to subjects who immediately faced the more sensitive questions?

Although the effect of question order has been studied extensively (e.g.: McFarland 1981; Barnes, Banahan, and Fish 1995; and many others), the impact of questions of different degrees of *intrusiveness* is underexplored. Hui (2007) found no statistically significant impact of the order of personal questions on individuals' propensity to answer them in an online shopping task. Moon (2000) studied how individuals' propensity to answer personal questions asked by a computer changed with “familiarity” (the time a participant had previously

spent working on that particular computer), but again question intrusiveness was not varied between conditions. Altering the order of intrusiveness of a set of questions is akin to asking subjects to comply with requests of different magnitude, and therefore comparable to the literature on “foot in the door” (FITD) (Freedman and Fraser 1966) or “door in the face” (DITF) techniques (Cialdini, Vincent, Lewis, Catalan, Wheeler, and Darby 1975). However, to our knowledge, no study has investigated the impact of such techniques on self-disclosure.²

The present research is related to three streams of marketing research. First, our studies contribute to the literature on survey design, and in particular to the stream of studies on the impact of contextual factors on “self-reports” (e.g. Schwarz and Julia Bienias 1990; Schwarz 1999). Second, we build upon the social psychology literature on self-disclosure (e.g. Altman and Taylor 1973; Mikulincer and Nachson, 1991; Derlega et al. 1993). This literature has investigated numerous drivers of self-disclosure, some of which we invoke in our comparative account of the propensity to reveal sensitive information. Third, we contribute to the literature on the relationship between privacy concern and willingness to divulge (e.g., Margulis 2003; Joinson, Woodley, and Reips 2007), and its relevance to marketing (Culnan and Armstrong 1999; White 2004). Inasmuch as our results can be extrapolated beyond the narrow domain of survey responses, they suggest that privacy concerns (exemplified by unwillingness to reveal sensitive information) are malleable to non-normative factors.

CONCEPTUAL BACKGROUND, HYPOTHESES, AND EMPIRICAL APPROACH

Human judgment and decision making is inherently comparative in nature. A wide range of research shows that people tend to judge stimuli and make decisions in a comparative fashion, and that they do so automatically, and without conscious awareness. Comparative judgments are especially likely when there is no objective basis for evaluation, which is likely the case for self-disclosure. How much of a net gain (if any) have you experienced by disclosing your personal information in order to find out your biological age on data-gathering sites such

as “realage.com”? When attributes are difficult to evaluate in the absolute, people naturally seek out points of comparison (Hsee et al. 1999). We consider two such points in our analysis: the effect of others’ disclosures, and the effect of the ordering of question sensitivity.

The effect of others’ disclosures

Research from a variety of literatures has shown that people are powerfully influenced by the behavior of those around them. Applied to self-disclosure, research and theorizing about herding (Devenow and Welch 1996) would suggest that if large numbers of people are revealing some kind of information, there is probably not great risk (and there may even be a benefit) to doing so oneself. Similarly, research on social norms finds that people adapt their behaviors to conform to that of those around them (Asch 1955, 1958; Sherif 1966; Jones 1984; Cialdini and Trost 1998; Krupka and Weber 2008), and infer injunctive norms (what one *should* do) from observations of descriptive norms (what people actually do: Bicchieri 2006). Applied to self-disclosure, these findings imply that when people are surrounded by others who are revealing intimate details of their lives, they may conform to the prevailing norm of divulgence. In addition, research focusing on self-disclosure has found that motives such as desire for social approval (Baumeister and Leary 1995) and reciprocity (Kenny 1994) promote disclosure. Observing other people’s willingness to answer intrusive questions, and in particular to admit to sensitive behaviors, may lead respondents to be less concerned about social disapproval, and to in turn, reciprocate disclosure. These theories suggest that we should expect higher admission rates to sensitive behaviors when subjects observe other people more frequently admitting to having engaged in sensitive behaviors, and, conversely, lower admission rates when people observe others either denying engaging in behaviors or refusing to answer the questions:

H1: Information about higher admission rates of engagement in sensitive behaviors by others will lead to increased admission rates among respondents.

We test Hypothesis 1 in Study 1A. This “herding” effect may arise because it alters the experience of responding affirmatively. Specifically, we hypothesize that people experience less discomfort in disclosing information when told that others tend to disclose. In two follow-up studies, we provide supporting evidence for this explanation (namely, that the feedback affects people’s expectations about the experience of divulging sensitive information: Study 1B) and rule out alternative explanations (namely, that the

² Freedman and Fraser (1966) and Furse, Stewart, and Rados (1981) investigated whether the magnitude of initial requests affects individuals’ propensity to participate in a survey, but not their disclosure of sensitive information. Reingen and Keman (1979) and Mowen and Cialdini (1980) manipulated the length of the survey subjects were requested to comply with, but not the intrusiveness of its questions.

manipulation alters the interpretation of the questions: Study 1C).

The effect of the ordering of question intrusiveness

Other people are only one point of comparison; previous experiences are another. Evidence from a variety of literatures suggests that disclosure may indeed be influenced by comparisons to a previous disclosure, or request for disclosure. From a psychophysics perspective, an intrusive question about engaging in a sensitive behavior may appear more [less] intrusive when contrasted with tamer [more sensitive] enquiries, which would in turn affect a respondent's propensity to admit to having engaged in the behavior (for an account of the relationships between psychophysics and embarrassment, see Latane 1981). Similarly, heuristics such as coherent arbitrariness (Ariely, Loewenstein, and Prelec, 2003) and comparative ignorance (Fox and Tversky 1995) suggest that a subject's likelihood of answering intrusive questions is affected by the contrast between the current and the previous questions in a survey. Marketing accounts of FITD (Freedman and Fraser 1966) or DITF techniques (Cialdini et al. 1975) may similarly predict that admission rates in a sensitive survey may differ because the order of intrusiveness of questions alters people's perceptions of the intrusiveness of the questions.

Specifically, previous research has identified factors that predict whether DITF or FITD will apply in a given situation (Dillard 1991). Tybout, Sternthal and Calder (1983)'s "availability hypothesis" predicts that an individual's compliance with a request depends on the *favorableness* of the issue-relevant information available in memory, where "issue-relevant information" refers to the behavior of either the requester or the requestee. A tame request is an example of favorable request-behavior; an individual's compliance with a request is an example of favorable own-behavior. Tybout et al. propose that FITD enhances compliance only when favorable own-behavior (in our context, complying by answering survey questions about engaging in behaviors of various sensitivity) is more readily available in the individual's memory than unfavorable request-behavior (in our context, the increasing intrusiveness of the questions). Critically, this occurs only if respondents are first asked to accept a substantial initial request (Seligman, Bush, and Kirsch 1976) and if the availability of request-behavior has been reduced by making noncontiguous requests (Freedman and Fraser 1966; Seligman, Bush, and Kirsch 1976). On the other hand, DITF should enhance compliance when favorable request-behavior information (in our context, the fact that the

intrusiveness of questions decreases through the survey) is more "available" than own-behavior information. Presenting successive questions within a single survey and by a single requester should make request-behavior more salient than own-behavior, and should therefore lead to a DITF, instead of an FITD, effect. We therefore hypothesized that:

H2: Respondents presented with questions in decreasing [increasing] order of intrusiveness will be more [less] likely to admit to having engaged in sensitive behaviors than respondents presented with questions in random order of intrusiveness.

We test Hypothesis 2 in Study 2A. We propose that this effect occurs by altering people's perceptions of the intrusiveness of the questions – specifically, we expect people to judge the questions about sensitive behaviors to be less intrusive when they are presented in a decreasing order of intrusiveness. Accordingly, a corollary hypothesis we test in Study 2A is that the effect of our manipulation should be more significant for the most intrusive questions (relative to tamer ones), while a corollary hypothesis we test in Study 2B is that the ordering of questions directly affects people's judgments of the intrusiveness of the inquiries. Finally, we investigate under which conditions the effect can be altered. In Study 2C, we test whether the effect can be nullified by priming concerns over intrusiveness and privacy at the outset of the study. In Study 2D, consistent with the "perceptual contrast" account of DITF and FITD dynamics (Shanab and O'Neil 1982; Cantrill and Seibold 1986) - under which the initial request acts as an "anchor" against which further requests are interpreted - we test whether the effect can be nullified by forcing participants to consider the relative intrusiveness of questions before the survey begins.

Although we believe that both herding (H1) and order effects (H2) operate through comparative mechanisms, there is reason to believe that the specifics might be subtly different. Presenting information on others' admission rates should naturally draw attention to the act of admitting; therefore, the mechanism for the herding effect pertains to this experience – when told that others tend to respond affirmatively, we predict that people anticipate less discomfort in responding affirmatively, in turn increasing admission rates. By contrast, the explanation of the effect of the ordering of question sensitivity pertains to people's perceptions of the intrusiveness of the questions, rather than the act of admitting itself.

Empirical Approach

All studies were online questionnaires in which participants were asked questions about a series of different behaviors. Between-subjects, we

manipulated a factor expected to affect comparative judgments (feedback on others' admissions in Studies 1A-C, and the intrusiveness order in which the questions were presented in Studies 2A-D). Since each study included multiple questions that participants answered in sequence, we analyzed responses using econometric methodologies for panel data. As the answers provided by participants were (depending on the study) either dichotomous or ordinal, we estimated random effects probit or ordered probit models. The random effect specification enables us to measure the effect of the treatment on the dependent measure (the subject's answers), controlling for the non-independence of observations by the same subject, and unobservable individual differences (e.g. in privacy sensitivity and desire to disclose). The Web Appendix reports additional methodological details.

In the primary experiments (1A and 2A), participants indicated how frequently (and, therefore, whether) they had engaged in the behaviors. Since most of the behaviors were of a sensitive nature, admitting to having engaged in them carried potential costs, whether subjective (e.g. embarrassment) or objective (e.g. incrimination), which, we posited, would create an obstacle to responding affirmatively. Therefore, our dependent measure in the primary studies is the propensity to respond affirmatively (hereafter referred to as "admissions"). In the follow-up studies, participants rated various aspects of the questions as a function of the experimental manipulations, such as their perceived intrusiveness or clarity.

To make admissions more directly relevant to marketers, we also asked participants to provide email addresses. To provide an incentive for participants to respond truthfully, we offered them the option of receiving "personalized results, including where [they] fall relative to others on the traits and attitudes the survey measures." Insofar as participants were interested in accurate feedback, this feature created an incentive for truthful responding. However, and importantly, by "admission" we do not necessarily refer to *truthful* admissions: our interest is not in the true underlying prevalence, but in the comparative nature of people's willingness to divulge sensitive information by openly admitting to having engaged in embarrassing, socially unappealing, and even illegal behaviors. In other words, we were not attempting to measure the true prevalence of these behaviors *per se*, but how incentives for truthful responding would interact with resistance to embarrassing or incriminating self-disclosure as a function of our experimental manipulations.

Missing answers. It was possible for participants to leave items blank. In the analyses reported in the main body of the manuscript, we treat

such non-responses as neither admissions nor denials. However, missing answers may signal a participant's unwillingness to answer a question, or may simply be due to attrition. Both scenarios are of interest to us. Willing refusal to answer a question implies the absence of an *explicit* affirmative admission. Hence, we also analyzed our data in a specification that treats missing answers as non-admissions. The results are equivalent to those obtained when ignoring non-responses, and are reported and discussed in the Web Appendix. Attrition, on the other hand, could generate survivor bias. Our results, however, are not altered by survivor bias, either because the number of participants who did not complete the survey did not differ across conditions (Study 1A), or because the differences do not affect our conclusions (Study 2A). We also present the related analysis in the Web appendix.

The Behaviors. The behaviors varied in sensitivity – from tame (e.g. *Failing to tip a waiter in a country in which tipping is customary*) to highly sensitive (e.g. *Having sex with the current partner of a friend*) and even illegal (e.g. *Cheating on one's tax return*). We assessed the perceived intrusiveness of questions about these behaviors in a pilot study in which an independent sample of 25 students at a North American university rated each question on a 4-point scale (*Not at all intrusive*, *Mildly intrusive*, *Intrusive*, and *Very intrusive*).

STUDY 1

Study 1A

Study 1A was a three condition between-subjects randomized experiment in which we manipulated the distribution of answers ostensibly supplied by other participants in the same survey; in reality, this was a fictional distribution. Participants were told that a relatively large proportion of previous respondents had responded affirmatively (High condition), or had responded with denials (Low condition), or had not responded at all (Missing condition).

Procedure. Participants were directed to the questionnaire by a link titled "Test your ethics" in the online version of the New York Times, and were randomly assigned to one of the three experimental conditions. Participants were told that they would be presented with descriptions of a series of behaviors and questions about them. They were also told that they would be informed, *after* answering each question, of the current distribution of other respondents' answers. Participants were then asked to provide their email address and to answer a series of demographic questions. Participants were presented with six pairs of questions; each pair pertained to a specific behavior and was presented on its own page.

In the first question of each pair, participants were asked to rate the ethicality of the behavior (*Not at all unethical, Somewhat unethical, Quite unethical, Extremely unethical, It depends, and Nothing to do with ethics*). In the second question of each pair, participants were asked to indicate how frequently, if ever, they had engaged in the behavior (*Never, Once or twice, Sometimes, Frequently*).

After answering each question, participants could observe the distributions of answers ostensibly given by previous respondents. The distribution showed the percentage of respondents who had admitted to having engaged in the behavior, had denied having engaged in the behavior, or had not answered the question. To increase the salience of this information, the distribution of answers was presented visually, in histogram format. In the “High” condition, the histograms depicted that a majority of other respondents had responded affirmatively. In the “Low” condition, the histograms depicted that a majority of other participants had denied having engaged in the behaviors. In the “Missing” condition, the histograms depicted that a majority of other respondents had left the questions blank. Although the admission rates were always either high or low within a given condition, the exact rates varied between behaviors within each condition to make the feedback credible (see Figure 1). The questions were presented in the same order across the conditions. All questions, except the first one, were picked from the set judged as very intrusive in the pilot study (Appendix A).

[Figure 1 about here]

Since we were interested in differences in affirmative admission rates (from now onwards: AARs), we hypothesized that participants in the High condition (who observed high AARs) would be more likely to report having engaged in the behaviors than subjects in the Low or Missing conditions (who observed low AARs).

Importantly, since the ostensible distribution of answers to a given question was shown only *after* the participant had answered the question (and could not go back to change his or her answer), the effect we tested was not the trivial impact of other people’s admission to a *given* behavior on the individual’s propensity to admit to that *same* behavior; but rather, whether the overall admission to sensitive behaviors would make the participant more likely to admit to *other*, also sensitive, behaviors. Therefore, our results are not due to mere imitation of other “respondents” exact responses to the same question.

Empirical approach. The dependent variable of primary interest was whether participants admitted to having engaged in a behavior. Since we were

interested in whether participants admitted to having engaged in a behavior as function of our manipulations, rather than their reported *frequency* of engagement, we collapsed the four frequency categories (*Never, Once or twice, Sometimes, Frequently*) into one dichotomous variable which we analyzed using a probit specification (0 = never engaged in the behavior; 1 = engaged in the behavior at least once). However, the results we present below are robust to the consideration of the original, 4-point ordinal dependent variable in an ordered probit specification.

Results. New York Times website visitors took the survey (N = 1,722; mean age = 40 years; 45% males; 82% Caucasian; males were slightly more represented in the High condition - significant at the 5% level; however, the results do not change when we control for gender or other demographics. All other demographic traits were similarly distributed between conditions).

Table 1, first set of columns, presents the results of the random effects probit specification. As hypothesized, AARs were significantly higher in the High condition relative to the Low and Missing conditions. The coefficients for both the Low condition dummy and the Missing condition dummy are negative and significant at the 5% level (the baseline condition is the High condition).³ Table 2 presents admissions rates to each individual question. Given the very different base rates of admissions between questions, we examine the mean percent difference in admissions rates between conditions, averaging over questions. Participants in the High condition were on average 27% more likely to admit to having engaged in the behaviors than participants in the Low condition ($t(948) = 3.74, p = 0.0001$), and 21% more likely to admit to having engaged in the behaviors than participants in the Missing condition ($t(942) = 2.99, p = 0.0014$).⁴ In contrast, there were no statistically significant differences between the AARs in the Low and the Missing conditions. Table 2 also shows how – except for the first question (*Have you ever bounced a check*), which was asked *before* the manipulation actually started – AARs in the High condition were always higher than the AARs in both other conditions, and statistically significantly so for two of the three last questions (also after applying a Bonferroni correction for

³ Regression results with demographics are equivalent for all studies and are available from the authors on request.

⁴ Responses to the first question were excluded from the results, as the ostensible distribution of answers to each question was shown to the participant *after* she had already answered it. The overall differences between conditions is weakened, yet remains significant, when including the answers to the first question.

multiple comparisons).

[Tables 1 and 2 about here]

Since the experimental treatment was applied sequentially and repetitively following each question, a corollary of Hypothesis 1 would predict that the impact of the manipulation would increase as participants answered successive questions. This was in fact the case. A version of the random effects probit with interaction confirms that interactions between the last questions in the survey and the Missing and Low conditions are negative and significant (Table 1, second set of columns). Furthermore, Figure 2 shows the *cumulative* distribution of AARs across conditions (excluding answers to the first question). In addition to illustrating the increasing departure of the High condition from the other two conditions, the figure highlights the equivalence of the AARs in the Low and Missing conditions.

[Figure 2 about here]

Across all conditions, the majority of participants provided email addresses. Overall, participants who provided email addresses were not only more susceptible to the manipulation than those who did not, but were also more likely to admit to having engaged in the behaviors. Participants were equally likely to complete the survey across conditions, and our results are robust to the provision of potentially identifying information. Our results are also robust to coding missing answers as non-admissions (participants in the High condition remain 19% more likely to admit to having engaged in the behaviors than participants in the Low and Missing conditions). Details about email and non-response analyses are presented in the Web Appendix.

Study 1B

Participants in Study 1A were more likely to admit to having engaged in sensitive behaviors when told that a relatively high proportion of previous respondents had made similar admissions. In Study 1B, we test a possible explanation for this effect: seeing that many other respondents had felt comfortable responding affirmatively may have affected respondents' expectations about the experience of responding affirmatively; specifically, it might make them anticipate less discomfort in responding affirmatively. We test this idea by measuring people's perceptions of how uncomfortable it would be to respond "yes" to each question, as a function of the manipulation used in Study 1A. We hypothesized that participants in the

High condition would expect *lower* discomfort in responding affirmatively relative to those in the Low condition.

The design was equivalent to the one employed in Study 1A except that, instead of indicating whether they had engaged in the behaviors, participants were asked: "How uncomfortable would it be for you to respond 'yes' to this question?" Participants responded on a 4-point scale labeled: *Not at all uncomfortable*, *Somewhat uncomfortable*, *Uncomfortable*, *Very uncomfortable*. The behaviors were the same as those used in Study 1A. As in Study 1A, *after* each answer about a given behavior, participants were shown a histogram that depicted the proportion of individuals who had ostensibly previously indicated that they had engaged in the behavior. We ran two conditions in Study 1B: the histograms were manipulated to depict either relatively high or relatively low AARs.

Results. Participants (N = 247; mean age = 34 years; 39% male; 83% Caucasian) were recruited online and randomly assigned to the High or Low conditions, and were given a small fixed payment at the end of the study. A random effects ordered probit model indicated, as hypothesized, that discomfort ratings were statistically significantly lower in the High condition relative to the Low condition (Table 3, first set of columns). Averaging across Questions 2 to 6, the mean reported discomfort in the Low condition was 1.90; in the High condition, it was 1.72 ($t(243) = 1.76$, $p = 0.0402$; as expected, discomfort ratings were not different between conditions for Question 1, which was asked *before* the feedback manipulation took place).

[Table 3 about here]

Alternative Explanations and Study 1C

Study 1A suggests that people are more likely to report that they have engaged in sensitive behaviors when they are lead to believe that others, too, have admitted to engaging in other sensitive behaviors. Study 1B suggests that seeing a majority of individuals willing to make such admissions makes it less uncomfortable for a person to do so himself. These findings suggest that herding behavior affects the propensity to disclose sensitive information. There are, however, a number of alternative and more mundane possible explanations for this effect.

First, in the High condition, seeing that a large proportion of other participants had responded affirmatively may have simply made affirmation a more available response. However, a closer look at the results of Study 1A suggests that this is unlikely. The histograms representing other respondents' ostensible AARs collapsed the response options into

three categories (*Never did, Did at least once, Refuse to answer*), while actual respondents used a different, 4-point scale (*Never, Once or twice, Sometimes, Frequently*). If the results were merely driven by “*Did at least once*” being a more available response, one would expect participants to be more likely to simply affirm that they had engaged in the behaviors *once*. Instead, our manipulation is also significant in an ordered probit specification of Study 1A where the DV is ordinal (from *Never* to *Frequently*) instead of dichotomous. In fact, the percentage of participants claiming to have engaged in a behavior *more than once* is larger in the High condition (10.20%) than in the Low and Missing conditions (6.99% and 4.38% respectively; Pearson $\chi^2(2) = 9.16, p = 0.0100$). This suggests that the *entire* distribution of reported frequencies shifts “to the right” in the High condition: participants do not simply admit to having engaged in more behaviors – they actually report *higher frequencies* of engagement.

Study 1C. A second alternative explanation is that the information about other people’s admissions affected people’s construal of the behaviors in question – consistent with previous research demonstrating how a survey’s design (e.g. question order, response options) can shape respondents’ interpretations of the questions posed therein (e.g., Schwarz and Scheuring 1988). In the High condition, participants may have inferred the behaviors to be broadly defined; seeing more people admitting to them may have lead participants to think of more instances in which they had engaged in sensitive behaviors. For example, seeing that a large proportion of previous respondents had admitted to “cheating on a partner,” participants may have interpreted this item broadly, to include a wide range of activities – from flirting with a person other than one’s relationship partner to having sexual intercourse with such a person. By contrast, in the Low or Missing conditions, seeing that few previous respondents had responded affirmatively, participants may have inferred the behaviors to be narrowly defined, resulting in lower AARs. However, a follow-up experiment (Study 1C) suggests that this explanation cannot account for our results.

The design was similar to Study 1A, except that participants were told that we were “not interested in whether you have engaged in the given behavior. Instead, we are interested in how you think other subjects interpret those behaviors. We will present you with two descriptions of each behavior, and ask you to choose the one you think best describes how other [subjects] interpret that behavior.” As in Study 1A, participants were first presented with a behavior (e.g., “While an adult, having sexual desires for a minor”). Then, they were asked to check which of two descriptions best

described how they thought other subjects would interpret the phrase. For each behavior, they were presented with two descriptions: a literal description (e.g., “As an adult, feeling desires of a sexual nature towards a minor”), and a broader one (e.g., “As an adult, finding a minor attractive”). *After* each answer, participants were shown histograms depicting the proportion of subjects who had indicated, in a previous survey, that they had engaged in the behavior.

Results. We recruited participants online (N = 126; mean age = 32 years; 47% male; 77% Caucasian; no significant differences between conditions). A random effects probit model shows no significant difference between the breadth of definitions chosen by participants in the High versus Low conditions (Table 3, third set of columns). This result suggests that the herding manipulation does not affect participants’ interpretation of question breadth.

STUDY 2

Study 2A

In Study 2A, we tested the effect of ordering questions along different gradients of intrusiveness (determined by ratings from the pilot study) on the propensity to respond affirmatively. As in Study 1A, participants first judged the ethicality of each behavior and then reported whether they had engaged in the behavior. The response options were the same as those used in Study 1A. The study was a 2x4 (between-subjects) x3 (within-subjects) randomized experiment. In the “Front” conditions, participants were asked to provide potentially identifying information (email address) at the beginning, of the questionnaire; in the “End” conditions, they were asked to provide this information at the end of the questionnaire. More importantly, we manipulated the order in which questions of different sensitivity were presented. In the “Increasing” conditions, the questions were presented in an increasing order of intrusiveness. In the “Decreasing” conditions, the order was reversed: participants first faced questions about the most sensitive behaviors; the questions became progressively tamer through the questionnaire. In the “Random” conditions, the questions were presented in a pseudo-random order of intrusiveness (i.e. the questions were placed in a jumbled order with respect to their intrusiveness). The Random condition was included to pinpoint whether the Decreasing condition facilitates admissions and/or whether the Increasing condition inhibits admissions. The “Sudden” conditions consisted of only tame questions, except for the *last three* questions, which were identical to the last three in the Increasing condition (and were therefore highly intrusive). The Sudden conditions served as an

alternative control, to test the propensity to admit to the most sensitive behaviors for participants who initially faced tame questions. Finally, within-subjects, we examined the propensity to admit to questions of different sensitivity (tame, moderate, and intrusive).

As in Study 1A, the dependent variable of interest was the propensity to respond affirmatively, estimated through a probit model. We tested whether this propensity depends on a) the order in which questions are presented with respect to their intrusiveness and b) the sensitivity of the questions. The results presented below are also robust to the consideration of the actual reported frequencies of engagement in an *ordered* probit specification.

Empirical approach. The empirical approach was equivalent to that used in Study 1A. To take into account the differences in question intrusiveness, we used the results of the pilot study to include categorical dummies in the regressions, representing the 10 tamest questions (e.g. “Have you littered in a public space?”), the 10 moderate questions (e.g. “While in a relationship, have you flirted with somebody other than your partner?”), and the 10 most intrusive questions (e.g. “Have you masturbated at work or in a public rest room?”). The complete list of questions is available in Appendix B.

Missing observations are of even greater importance in Study 2A, because the order in which questions were presented varied between conditions, and order effects could, in turn, interact with the participants’ propensity to leave questions blank (for instance, because the participant chose to abandon the questionnaire altogether). In the Web Appendix, we analyze missing observations relative to their placement within the questionnaire. The results presented there, however, are equivalent – for the intrusive questions – to those presented in this manuscript: our main findings are robust to the consideration of missing answers as non-admissions and are not determined by survivor bias.

Results. Readers of the online edition of the New York Times participated in the study (N = 2,310; mean age = 38 years; 65% male; 88% Caucasian; no significant demographic differences between conditions). Participants were significantly more likely to provide email addresses in the Front conditions relative to the Back conditions; however, the point at which they were asked to supply their email address did not interact with the question order manipulation. We therefore collapse across the Front and End conditions for the rest of the analysis, and refer to Increasing, Decreasing and Random conditions in the singular.

We begin by focusing on the Increasing, Decreasing and Random conditions (which, unlike the Sudden condition, are comparable because they

contained the same questions, albeit in different orders). Table 4, first set of columns, presents the results of the random effects probit specification. As hypothesized, across all questions (tame, moderate, and intrusive) participants in the Increasing conditions were less likely to admit to behaviors than those in the Decreasing condition; there was no such difference between the Random and Decreasing conditions. Table 4, second set of columns, includes the interaction terms. It shows that participants in the Decreasing condition were significantly more likely to admit to the most *sensitive* behaviors than participants in both the Increasing and Random conditions. This finding confirms that our manipulation is particularly significant for questions associated with the most sensitive behaviors. Furthermore, in both specifications, AARs were lower for sensitive and moderately intrusive questions relative to the tame questions.

Table 5 presents admissions rates for each individual question, ordered from most intrusive to least intrusive. Note that the Decreasing condition exhibits high non-admission rates for the first question (which is very sensitive). While the low AARs for that question are not significantly different across conditions, significant differences emerge in subsequent questions, with participants in the Decreasing condition being more likely to admit to the subsequent (still sensitive, but less so) behaviors. This pattern is consistent with the DITF dynamics highlighted earlier in the manuscript. (As a caveat, we note that the analysis of AARs to individual questions is confounded by idiosyncrasies specific to each question, whereas categorical dummies such as “tame questions,” which we used in the regression, provide a more robust analysis.) Averaging the percent differences in AARs across questions, participants in the Increasing condition were 19% less likely than participants in the Decreasing condition to admit to having engaged in the behaviors ($t(917) = 6.64, p < 0.0005$), and 18% less likely than those in the Random condition ($t(939) = -6.58, p < 0.0005$). As for the most intrusive questions, participants in the Increasing condition were 20% *less* likely to admit to the sensitive behaviors ($t(980) = 3.92, p < 0.0005$) than participants in the Random conditions; while participants in the Decreasing condition were 15% *more* likely to admit to the most sensitive behaviors ($t(969) = 2.79, p = 0.0027$) than participants in the Random conditions. Furthermore, participants in the Increasing condition were 51% less likely than those in the Decreasing condition to admit to having engaged in the 10 most sensitive behaviors ($t(967) = 6.84, p < 0.0005$). By contrast, the differences in AARs between conditions are much less dramatic for the less intrusive questions.

[Tables 4, 5 about here]

Figure 3a and 3b display the *cumulative* AARs, question by question, across the three main conditions. Across the entire survey, AARs in the Increasing condition lag behind those of the other conditions (Figure 3a); this difference in AARs is particularly pronounced for the most intrusive questions (Figure 3b). Additionally, for the most intrusive questions, AARs are higher in the Decreasing relative to all other conditions.

As noted above, our main results are also robust to the consideration of missing observations as non-admissions and are not altered by survivor bias: overall AARs are lower in the Increasing conditions, and AARs to the most intrusive questions are higher in the Decreasing condition (details available in the Web Appendix).

[Figures 3a and 3b about here]

Sudden condition. The sudden condition included only tame questions, except for its last three questions, which were the same questions as the last three in the Increasing condition (i.e., the three most intrusive questions). The mean admission rate to the three intrusive questions was the same as that of the Increasing condition (0.24 vs. 0.24; t-test: $p > 0.6$), implying that – similar to the Increasing condition – participants in the Sudden condition were significantly less likely to admit to having engaged in the three most sensitive behaviors relative to the Decreasing (t-test $p < 0.0005$) and Random (t-test $p < 0.0355$) conditions. These results confirm that, regardless of whether the shift in question intrusiveness is gradual or sudden, answering intrusive questions after non-intrusive ones inhibits admission.

Study 2B

Study 2A showed that people's willingness to admit to having engaged in sensitive behaviors depends on the intrusiveness of previous such inquiries: participants who faced questions of increasing [decreasing] sensitivity were *less* [more] likely to admit to the most intrusive questions than subjects presented with questions in random order of intrusiveness. However, whereas the manipulation in Study 1A, by emphasizing the AARs from other respondents, draws attention to the act of admitting, the manipulation in Study 2A does not. In other words, whereas the herding manipulation affects people's anticipated discomfort in responding affirmatively (Study 1B), we hypothesized that, in Study 2A, the question ordering manipulation would affect judgments of the intrusiveness of the questions. This would imply, consistent with AARs, that people

would judge questions to be less intrusive when they are presented in a decreasing order of intrusiveness, relative to when they are presented in an increasing order of intrusiveness. We test this idea in Study 2B.

Procedure. Study 2B was a two condition between-subjects design similar to Study 2A. Participants were presented with a series of questions about different behaviors, ranging from tame to intrusive. To test Study 2A's robustness to a different sequence of questions, and to further reduce the risk of survivor bias, we shortened the number of items from 30 to 6 (the three tamest and the three most intrusive questions in Study 2A). In the Decreasing condition, the questions were presented in a decreasing order of intrusiveness; in the Increasing condition, the questions were presented in an increasing order of intrusiveness. For each question, participants were asked to "rate how intrusive (if at all) the question is," using a 4-point response scale (*Not at all intrusive, Mildly intrusive, Intrusive, and Very intrusive*).

Results. Participants ($N = 133$; mean age = 33 years; 42% male; 33% Caucasian; no significant differences between conditions) were recruited online and randomly assigned to one of the two conditions. Table 6 presents the results of a random effects ordered probit model: as hypothesized, participants in the Increasing condition judged the questions as more intrusive than those in the Decreasing condition. The mean intrusiveness rating was 2.09 in the Decreasing condition and 2.44 in the Increasing condition. (We also ran an additional specification testing the interaction between question intrusiveness and the experimental manipulation: The manipulation remains significant, but the interaction is not, as the mean intrusiveness rating is higher in the Increasing condition both for the three least intrusive questions and the three most intrusive ones.)

[Table 6 about here]

Study 2C

Taken together, Studies 2A and B suggest that individuals are less likely to admit to having engaged in sensitive behaviors when the questions are presented in escalating order of sensitivity, altering their perceived intrusiveness. A possible interpretation of these results, suggested by some of the literature explored earlier in the paper, is that the differential propensity to disclose is linked to people's malleable concerns about the privacy of their personal information. In other words, it is possible that by altering the order of questions (and therefore their perceived intrusiveness), our manipulation affects privacy concerns. If that were

the case, priming participants with considerations of privacy before the survey is administered should have a similar effect to changing the order of question intrusiveness from decreasing to increasing – i.e., making them ‘clam up’ and admit to fewer sensitive behaviors. We test this prediction in Study 2C.

Study 2C was a 2x2 between-subjects design in which we manipulated whether participants were cued to think of privacy from the outset (privacy cue vs. no cue), along with the order in which the questions were presented (Increasing vs. Decreasing order of intrusiveness). We hypothesized that the privacy cue would lead participants in either condition to admit less, regardless of the order of the questions. We also hypothesized that, holding the privacy cue manipulation constant, the results of Study 2A would be replicated (i.e. AARs would be higher in the Decreasing condition). More importantly, we hypothesized that once cued to think about privacy, participants in the Decreasing condition would be no more likely to respond affirmatively relative to participants in the Increasing condition whose privacy concern had not been roused. In other words, we hypothesized that the impact of changing the order of the questions from decreasing to increasing on the propensity to admit would be similar to the impact of cueing participants to think about privacy concerns. We did not need to hypothesize an interaction between the privacy cue and order of questions: participants in both the Increasing and Decreasing condition faced extremely intrusive questions that were likely, by themselves, to trigger concerns; hence, the privacy cue manipulation did not elicit a previously non-existent concern, but would rather heighten the concerns already aroused in either condition.

Procedure. All participants were asked to complete two surveys. In the first survey, participants had to complete a “photo identification task.” Participants were asked to either “Phind the phishing emails” (privacy conditions) or to “Find the endangered fish” (control conditions). In the privacy conditions, participants were given definitions of phishing (emails fraudulently claiming to be reputable that attempt to lure recipients into divulging sensitive information) and spam taken from Wikipedia. In the control conditions, participants were given a definition of endangered species, also from Wikipedia. To reinforce the manipulation, on the subsequent pages, participants were asked to define the term “phishing” or “endangered species” (depending on the condition). On the following six pages, participants were asked to categorize various images presented to them. In the privacy conditions, the images were screen shots of email messages; participants indicated whether each image constituted phishing or spam. In the control conditions, the

images were photos of fishes, and participants indicated whether each species was endangered or not. After finishing the first survey, participants clicked a button to begin the second survey. At that point, participants were randomly assigned to one of two versions of the survey used in Study 2A – Increasing or Decreasing.

Results. Participants were recruited online from a pool of New York Times readers; each was randomly assigned to one of four experimental conditions (N = 161; mean age = 47 years; 51% male; 88% Caucasian. There were no significant demographic differences between conditions, except for an over-representation of males in the Decreasing condition, significant at the 10% level; the results presented below do not change when controlling for demographics).

A random effects probit model revealed that both the privacy cue manipulation and the Increasing order manipulation significantly decreased participants’ propensity to admit (Table 7, first set of columns; the interaction, as discussed, is not significant). Table 7, second set of columns, presents the results of a further specification in which we contrasted only two conditions: the Increasing condition *without* privacy cueing, and the Decreasing condition *with* privacy cueing. The higher propensity to admit elicited by the Decreasing manipulation disappears: the coefficient for the dummy representing the order of questions manipulation (*Increasing condition*) is no longer significantly different from zero ($p > 0.6$). These results are also reflected in the mean AARs across conditions. The mean admission rate is highest in the Decreasing condition with no privacy cue (0.43) and lowest in the Increasing condition with privacy cue (0.33), but it is virtually the same for the Decreasing condition with privacy cue (0.38) and the Increasing condition without (0.37).

Beyond replicating the results of Study 2A, Study 2C shows that cueing people to think about privacy from the outset of the experiment decreases their propensity to admit: once cued to think about privacy, participants in the Decreasing condition are no longer more likely to respond affirmatively than participants in the Increasing condition whose privacy concerns had not been roused.

Study 2D

Study 2D bridged the designs and goals of Studies 2B and 2C. It tested whether Study 2A’s results would also disappear (as in Study 2C) when participants are induced to think about, and rank, the relative intrusiveness of a set of personal questions (as in Study 2B), *before* answering questions about their own behavior. Such a finding would support the interpretation of Study 2A’s results as an outcome of

“perceptual contrast” DITF/FITD dynamics (Shanab and O’Neil 1982; Cantrill and Seibold 1986): the perception of the intrusiveness of personal questions changes with the order in which those questions are presented, which in turns impacts participants’ propensity to admit to the behaviors. Hence, if participants are primed at the onset to think about and contrast questions of varying intrusiveness, the differential impact of the order of questions on propensity to admit should be nullified.

Study 2D was a 2x2 between-subjects design in which we manipulated whether participants were cued to compare the *intrusiveness*, or instead the *wordiness*, of a set of questions about personal behaviors from the outset, along with the order in which a different set of questions were presented to them (Increasing vs. Decreasing order of intrusiveness). We hypothesized that, in the conditions in which participants were asked to rate the wordiness of questions about behaviors, we would replicate the findings of Study 2A (that is, question order would impact propensity to admit, with higher AARs in the Decreasing condition); whereas, in the conditions in which participants were asked to rate the intrusiveness of questions about behaviors, we would no longer find any impact of question order on propensity to admit.

Procedure. All participants were asked to complete two surveys. In the first survey, participants were asked to rate the “intrusiveness” of 18 questions about various behaviors (intrusiveness conditions: “How intrusive, if at all, is each of the following questions? Note: please rate the intrusiveness, independently of whether or not the behavior in question applies to you.”) or the “wordiness” (control conditions: “How wordy, if at all, is each of the following questions? Note: please rate the wordiness, independently of whether or not the behavior in question applies to you.”). The behaviors ranged in sensitivity, from low to high (as measured in a pilot survey). Questions about the behaviors were presented in pseudorandom order of sensitivity. Thereafter, participants continued onto the second survey, which was equivalent to the survey used in Study 2A: participants were presented with six behaviors (three sensitive, three non sensitive) and asked how frequently, if ever, they had engaged in them. Depending on the condition, the behaviors were presented in either Increasing or Decreasing order of intrusiveness. There was no overlap between the set of behaviors included in the first survey and those included in the second survey.

Results. Participants were recruited online and randomly assigned to one of four conditions (N = 276; mean age = 35; 37% males; 81% Caucasian; there were no significant demographic differences between conditions). Participants were given a small

fixed payment at the end of the study. A random effects probit model shows that in the Control conditions (in which participants were primed to think about the wordiness of questions about different behaviors) the results of Study 2A are replicated: the Increasing order manipulation significantly decreases participants’ propensity to admit (see Table 7, third set of columns). Instead, in the Intrusiveness conditions (in which participants were primed to think about the intrusiveness of questions about different behaviors) the order of questions no longer differentially impacts participants’ propensity to admit (see Table 7, fourth - and last - set of columns).

Beyond replicating the results of Study 2A, Study 2D suggests that cueing people to think about the intrusiveness of questions from the outset of the experiment nullifies the impact of order of questions on propensity to admit.

Analysis

Study 2A showed that people are less likely to admit to having engaged in sensitive behaviors when the questions are presented in escalating order of sensitivity, relative to a condition where questions are presented in descending order. Studies 2B and 2C – as well as the fact that the reduction in propensity to admit is more pronounced for the most intrusive questions - shed light on the process underlying this effect. Study 2B shows that question order affects perceptions of the intrusiveness of the questions – when the questions are presented in decreasing order of intrusiveness, they are judged to be less intrusive relative to when they are presented in an increasing order. Moreover, this difference in intrusiveness ratings is not a mere reflection of the fact that participants in the Decreasing condition were also more likely to admit to having engaged in the behavior (i.e., it cannot be a simple by-product of the increased tendency to respond affirmatively in the descending condition) because participants judged the intrusiveness of the questions *without indicating whether they had engaged in the behaviors*. Study 2C and 2D provide further evidence of the explanation for Study 2A’s results, by showing that when participants in the Decreasing condition are cued to think about privacy concern (Study 2C), their disclosure levels are similar to those in the Increasing condition *not* cued to think about privacy; and when they are cued to think about the intrusiveness of various questions before reporting their actual behavior, their propensity to admit is similar to those in the Increasing condition (Study 2D).

DISCUSSION

To the delight of marketers, new technologies have facilitated the acquisition, storage,

and integration of consumers' personal information on a mass scale. These technological advances have also, however, made it increasingly difficult for consumers to navigate issues of self-disclosure – that is, to choose an 'optimal' balance between information protection and information sharing in different situations. This can be a problem for marketers, too: the deluge of requests for personal data may either lead consumers to reveal more or, to clam up and become less willing to disclose.

In this paper, we provided evidence that the inherently comparative nature of human judgment and decision-making plays out in the way individuals decide to reveal personal, potentially embarrassing and even incriminating, information. Specifically, we showed that judgments of, and responses to, requests for sensitive information depend crucially on two points of comparison: the judgments and responses of other people, and the order in which questions of different sensitivities are presented. We also found that our manipulations seem to affect the feeling of discomfort or intrusiveness associated with the surveys, but not the perceived clarity of its questions. In combination, therefore, our studies support the hypothesis that people's decisions to disclose sensitive information are comparative in nature.

It is important to note that our study focused on determining how comparative valuations affect an individual's propensity to report to others certain information about herself. However, our studies were not designed to establish 'true' prevalence estimates of the behaviors in question, and were limited to a specific type of information that consumers may feel uncomfortable divulging (engagement in embarrassing or sensitive behaviors) as opposed to other types of information, such as Social Security numbers. Hence, we cannot tell whether, beyond affecting people's propensity to admit to behaviors they had engaged in, our manipulations may have also caused people to admit to having engaged in behaviors in which they had actually *never* engaged. The fact that our results tend to be strongest for intrusive items, such as having sexual desires for a minor, seems to weigh against such an effect. Furthermore, in the High admission condition in Study 1A, 27.2% of participants claimed to have had sexual desires for a minor, 11.7% claimed to have had sex with the current husband, wife, or partner of a friend, and 31.6% claimed to have fantasized about having violent, non-consensual sex with someone. The percentages virtually match those provided in the Decreasing condition in Study 2A (24.6%, 11.4%, and 30.2%, respectively). Either two completely different treatments led participants to lie to the same degree; or, in fact, participants in the High and Decreasing conditions were more comfortable responding affirmatively to behaviors in which they

had engaged relative to other participants in the respective studies.

Marketing researchers and professionals frequently use online surveys, games, and quizzes aimed at inferring individuals' personal information. Our results highlight some challenges in choosing the structure and timing of personal inquiries, in a context in which consumers are influenced by multiple requests for personal information and surrounded by streams of information about others. The present research has implications for the design of marketing surveys, especially those involving intrusive questions and sensitive behaviors. Instructions on survey design generally suggest that researchers open their questionnaires with general, milder questions.⁵ In contrast, and somewhat at odds with the guidance from this literature, our results suggest that starting with milder questions only to get to more intrusive questions may actually elicit lower overall willingness to divulge. Furthermore, our studies also suggest that when survey takers are provided (or otherwise get access to) information about the occurrence of certain behaviors, said information can significantly affect their propensity to reveal personal and sensitive information about themselves.

Perhaps the most important implications of our results are, however, those for consumer welfare. New information technologies have enhanced consumers' ability to communicate and interact, but have also raised novel and troubling issues about the privacy and security of personal data. These considerations have generated renewed interest in the trade-offs between privacy and (for instance) personalization, which has been described as the future of interactive marketing (Deighton 1996). Implicit in much of the literature dealing with privacy trade-offs is the assumption that consumers are rationally informed agents with stable preferences for self-disclosure and privacy (Posner 1978). Our results, however, suggest a different story. Self-disclosure seems to be affected by information about others' divulgements and the mere order in which sensitive enquiries are presented. Inasmuch as our privacy account is valid, our results suggest that privacy concerns are also malleable to the influence of comparative judgments. If privacy preferences are

⁵ See, for instance, “[i]deally, the early questions in a survey should be easy and pleasant to answer [...] whenever possible leave difficult or sensitive questions until near the end of your survey” from Creative Research Systems at <http://www.surveysystem.com/sdesign.htm>, and “[f]irst questions should be relevant and easy... [p]otentially objectionable questions are placed near the end,” from Penn State Survey Research Center at www.ssri.psu.edu/survey/qd.ppt. See also Payne (1951, p. 34).

thus instable, doubts arise about which behavior represents the “true” desired level of information protection and revelation, and, more importantly, whether consumers can make self-interested decisions with respect to their data when interacting with increasing complex information technologies – decisions, in other words, that they do not stand to later regret.

REFERENCES

- Irwin Altman and Dalmas Taylor (1973). *Social Penetration: The Development Of Interpersonal Relationships*. New York: Holt, Rinehart & Winston.
- Dan Ariely, George Loewenstein, and Drazen Prelec (2003). “Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences,” *The Quarterly Journal Of Economics*, 118(1), 73-105.
- Solomon E. Asch (1955). “Opinions and Social Pressure,” *Scientific American*, 193(5), 31-35.
- Solomon E. Asch (1958). “Effects Of Group Pressure Upon The Modification and Distortion Of Judgments,” in: Eleanor E. Maccoby, Theodore M. Newcomb, and Eugene L. Hartley, Eds, *Readings In Social Psychology*. New York: Holt, Rinehart, and Winston, 174–183.
- James H. Barnes, Benjamin F. Banahan III, and Kelly E. Fish (1995). “The Response Effect Of Question Order In Computer-Administered Questioning In The Social Sciences,” *Social Science Computer Review*, 13(1): 47 - 53.
- Roy F. Baumeister and Mark R. Leary (1995). “The Need To Belong: Desire For Interpersonal Attachment As A Fundamental Human Motivation,” *Psychological Bulletin*, 117, 497-529.
- Christina Bicchieri (2006). *The Grammar Of Society: The Nature and Dynamics Of Social Norms*, Cambridge University Press.
- Robert C. Blattberg and John Deighton (1991). “Interactive Marketing: Exploiting The Age Of Addressability,” *Sloan Management Review*, 33(1), 5–14.
- James G. Cantrill and David R. Seibold (1986). “The Perceptual Contrast Explanation of Sequential Request Strategy Effectiveness,” *Human Communication Research*, 13(2), 253-267.
- Robert B. Cialdini, Joyce E. Vincent, Stephen K. Lewis, Jose Catalan, Diane Wheeler, and Betty Lee Darby (1975). “Reciprocal Concessions Procedure For Inducing Compliance: The Door-In-The-Face Technique,” *Journal Of Personality and Social Psychology*, 31, 206-215.
- Robert B. Cialdini and Melanie R. Trost (1998). “Social Influence: Social Norms, Conformity, and Compliance,” In: Daniel T. Gilbert, Susan T. Fiske, Gardner Lindzey, Eds, *The Handbook Of Social Psychology*, 151-192.
- Mary Culnan and Pamela K. Armstrong (1999). “Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation,” *Organization Science*, 10 (1), 104-15.
- John Deighton (1996). “The Future Of Interactive Marketing,” *Harvard Business Review*, 74, 4-16.
- Valerian J. Derlega, Sandra Metts, Sandra Petronio, and Stephen T. Margulis (1993). *Self-Disclosure*. Newbury Park, CA: Sage.
- Andrea Devenow and Ivo Welch (1996). “Rational Herding In Financial Economics,” *European Economic Review*, 40, 603-615.
- James P. Dillard (1991). “The Current Status Of Research On Sequential-Request Compliance Techniques,” *Personality and Social Psychology Bulletin*, 17(3), 283-288.
- Craig R. Fox and Amos Tversky (1995). “Ambiguity Aversion and Comparative Ignorance,” *Quarterly Journal Of Economics*, 110(3), 585-603.
- Jonathan L. Freedman and Scott C. Fraser (1966). “Compliance Without Pressure: The Foot-In-The-Door Technique,” *Journal Of Personality and Social Psychology*, 4, 195-202.
- David H. Furse, David W. Stewart, David L. Rados (1981). “Effects of Foot-in-the-Door, Cash Incentives, and Followups on Survey Response,” *Journal of Marketing Research*, 18(4), 473-478.
- Chris K. Hsee, George Loewenstein, Sally Blount, and Max Bazerman (1999). “Preference Reversals Between Joint and Separate Evaluations Of Options: A Review and Theoretical Analysis.” *Psychological Bulletin*, 125, 576-590.
- Kai-Lung Hui (2007). “Consumer Disclosure: The Effects Of Company Information Presentation and Question Sequence,” Hong Kong University Of Science and Technology, Unpublished Manuscript.
- David A. Kenny (1994). “Using The Social Relations Model To Understand Relationships.” In R. Erber and R. Gilmour, Eds, *Theoretical Frameworks For Personal Relationships*. Hillsdale, NJ: Erlbaum, 111-127.
- Erin L. Krupka and Roberto A. Weber (2008). “Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?,” *IZA Discussion Paper* No. 3860.
- Leslie John, Alessandro Acquisti, and George Loewenstein (2011). “Strangers on a Plane: Context-dependent Willingness to Divulge Sensitive Information,” *Journal of Consumer Research*, 37, 858-873.

- Adam Joinson, Alan Woodley, and Ulf-Dietrich Reips (2007). "Personalization, Authentication and Self-Disclosure In Self-Administered Internet Surveys," *Computers In Human Behavior*, 23, 275-285.
- Stephen R. G. Jones (1984). *The Economics Of Conformism*. Oxford and New York: Blackwell.
- Bibb Latane (1981), "The Psychology Of Social Impact," *American Psychologist*, 36(4), 343-356.
- Stephen T. Margulis (2003). "Privacy As A Social Issue and Behavioral Concept," *Journal Of Social Issues*, 59(2), 243-261.
- Sam McFarland (1981). "Effects Of Question Order On Survey Responses," *Public Opinion Quarterly*, 45, 208-215.
- Mario Mikulincer and Orma Nachson (1991). "Attachment Styles and Patterns Of Self-Disclosure," *Journal Of Personality and Social Psychology*, 61(2), 321-331.
- John C. Mowen and Robert B. Cialdini (1980). "On Implementing the Door-in-the-Face Compliance Technique in a Business Context," *Journal of Marketing Research*, 17(2), pp. 253-258.
- Youngme Moon (2000). "Intimate Exchanges: Using Computers To Elicit Self-Disclosure From Consumers," *The Journal Of Consumer Research*, 26(4), 323-339.
- Anthony Ong and David Weiss (2000). "The Impact Of Anonymity On Responses To Sensitive Questions," *Social Psychology*, 30(8), 1691-1708.
- Stanley Payne (1951). *The Art Of Asking Questions*. Princeton University Press, 10th Printing, 1973.
- Richard Posner (1978). "An economic theory of privacy," *Regulation*, May-June, 19-26.
- Peter Reingen and Jerome B. Kernan (1977). "Compliance with an Interview Request: A Foot-in-the-Door, Self-Perception Interpretation," *Journal of Marketing Research*, 14, 365-369.
- James E. Reinmuth and Michael D. Geurts (1975). "The Collection of Sensitive Information Using a Two-Stage, Randomized response Model," *Journal of Marketing Research*, 12, 402-407.
- Norbert Schwarz (1999). "Self-Reports: How The Questions Shape The Answers," *American Psychologist*, 54, 93-105.
- Norbert Schwarz and Julia Bienias (1990). "What Mediates The Impact Of Response Alternatives On Frequency Reports Of Mundane Behaviors?," *Applied Cognitive Psychology*, 4(1), 61-72.
- Norbert Schwarz and Bettina Scheuring (1988). "Judgments Of Relationship Satisfaction: Inter- and Intraindividual Comparisons As A Function Of Questionnaire Structure," *European Journal Of Social Psychology*, 18(6), 485-496.
- Clive Seligman, Malcolm Bush, and Kenneth Kirsch (1976). "Relationship Between Compliance In The Foot-In-The-Door Paradigm And Size Of First Request," *Journal of Personality and Social Psychology*, 33(5), 517-520.
- Mitri E. Shanab and Pamela O'Neil (1982). "The Effects Of Self-Perception And Perceptual Contrast Upon Compliance With Socially Undesirable Requests," *Bulletin of the Psychonomic Society*, 19, 279-281.
- Muzar Sherif (1966). *The Psychology Of Social Norms*. Harper & Row.
- Roger Tourangeau and Yan Ting (2007). "Sensitive Questions in Surveys," *Psychological Bulletin*, 133(5), 859-883.
- Alice M. Tybout, Brian Sternthal, and Bobby J. Calder (1983). "Information Availability As A Determinant Of Multiple Request Effectiveness," *Journal of Marketing Research*, 20(3), 280-290.
- Tiffany Barnett White (2004). "Disclosure and Disclosure Avoidance: A Motivational Framework," *Journal Of Consumer Psychology*, 14(1/2), 41-51.

Table 1 – Study 1A: Random effects probit estimates

	I (DV: Engaged in Behavior)		II (DV: Engaged in Behavior)	
	<i>Coefficient</i>	<i>p</i>	<i>Coefficient</i>	<i>p</i>
Constant	.0964245	0.026	-.0588783	0.327
<i>High Condition</i>				
Low Condition	-.107732	0.018	.1907377	0.024
Missing Condition	-.0974721	0.032	.0673135	0.424
<i>Question 1 (Bouncing check)</i>				
Question 2 (Cheating on tax return)	-.9451133	0.000	-.8125121	0.000
Question 3 (False insurance claim)	-1.467787	0.000	-1.342974	0.000
Question 4 (Desire for minor)	-.6812493	0.000	-.4382892	0.000
Question 5 (Cheating on partner)	-1.265129	0.000	-1.113772	0.000
Question 6 (Fantasizing about non consensual sex)	-.5596698	0.000	-.2796096	0.001
<i>Interaction terms</i>				
Low*Question 2			-.3168739	0.011
Low*Question 3			-.2844941	0.046
Low*Question 4			-.4395456	0.000
Low*Question 5			-.3107992	0.021
Low*Question 6			-.4955467	0.000
Missing*Question 2			-.0837896	0.495
Missing*Question 3			-.0916801	0.515
Missing*Question 4			-.2937041	0.014
Missing*Question 5			-.1453321	0.275
Missing*Question 6			-.3524286	0.003
	<i>Prob > χ^2 = 0.0000</i>		<i>Prob > χ^2 = 0.0000</i>	
	<i>n = 1,538</i>		<i>n = 1,474</i>	

Table 2 – Study 1A: Percentages of participants admitting to having engaged in various behaviors

	Never	Once or more	No answer	<i>p</i>
<i>Bouncing a check</i>				
High condition	45.72	41.88	12.39	0.08
Low condition	40.78	49.65	9.57	
Missing condition	44.44	44.96	10.6	
<i>Cheating on one's tax return</i>				
High condition	68.94	18.32	12.74	0.386
Low condition	73.94	15.96	10.11	
Missing condition	69.74	17.95	12.31	
<i>Making a false or even somewhat inflated insurance claim</i>				
High condition	78.18	8.38	13.44	0.551
Low condition	81.74	7.27	10.99	
Missing condition	78.29	8.03	13.68	
<i>While an adult, having sexual desires for a minor</i>				
High condition	57.07	27.23	15.71	0.002**
Low condition	66.67	21.45	11.88	
Missing condition	63.42	21.37	15.21	
<i>Having sex with the current husband, wife, or partner of a friend</i>				
High condition	72.6	11.69	15.71	0.298
Low condition	76.6	9.93	13.48	
Missing condition	74.7	10.6	14.7	
<i>Fantasizing about having violent, non-consensual sex with someone</i>				
High condition	52.36	31.59	16.06	<0.0005***
Low condition	62.23	23.58	14.18	
Missing condition	59.83	23.59	16.58	

Note: Last column presents Pearson $\chi^2(1)$ p-values for relationship between condition (comparing high to low and missing) and admissions rates (comparing “once or more” to “never”) only for participants who provided an answer. Significance levels include Bonferroni correction for n=6; *: $p < .10$; **: $p < .05$; ***: $p < .01$. Significance levels are comparable for Pearson χ^2 calculated over the three conditions separately.

Table 3 – Studies 1B and 1C: Random effects ordered probit estimates

	Study 1B		Study 1C	
	DV:		DV:	
	Discomfort		Clarity	
	<i>Coefficient</i>	<i>p</i>	<i>Coefficient</i>	<i>p</i>
Constant	.226441	0.086	-2.276073	0.000
<i>Low Condition</i>				
High Condition	-.3005449	0.045	.060797	0.809
<i>Question 1</i> <i>(Bouncing check)</i>				
Question 2 (Cheating on tax return)	-.071444	0.544	.475004	0.110
Question 3 (False insurance claim)	-.5091979	0.000	.0087372	0.979
Question 4 (Desire for minor)	.6122745	0.000	1.104345	0.000
Question 5 (Cheating on partner)	.351803	0.002	.585664	0.048
Question 6 (Fantasizing about non consensual sex)	.6326698	0.000	.9916652	0.001
	<i>Prob > χ^2</i>	<i>= 0.0000</i>	<i>Prob > χ^2</i>	<i>=</i>
			<i>0.0002</i>	
		<i>n = 247</i>	<i>n = 121</i>	

Table 4 – Study 2A: Random effects probit estimates

	I		II	
	DV: Engaged in Behavior		DV: Engaged in Behavior	
	<i>Coefficient</i>	<i>p</i>	<i>Coefficient</i>	<i>p</i>
Constant	.3701102	0.000	.3393673	0.000
<i>Decreasing Condition</i>				
Increasing Condition	-.1596418	0.000	-.1173054	0.000
Random Condition	-.0119736	0.614	.0348463	0.274
<i>Tame Questions</i>				
Moderate Questions	-.8508995	0.000	-.8373933	0.000
Intrusive Questions	-.9094224	0.000	-.831757	0.000
<i>Interaction terms</i>				
Increasing*Moderate			-.0350908	0.348
Increasing*Intrusive			-.0935159	0.013
Random*Moderate			-.0041377	0.912
Random*Intrusive			-.1368417	0.002
	<i>Prob > χ^2 = 0.0000</i>		<i>Prob > χ^2 = 0.0000</i>	
	<i>n=1,581</i>		<i>n=1,581</i>	

Table 5 – Study 2A: Percentages of participants admitting to having engaged in various behaviors

	Decreasing	Random	Increasing	<i>p</i>
<i>Had sex with the current husband, wife, or partner of a friend?</i>	13.24	9.84	11.62	0.242
<i>Masturbated at work or in a public restroom?</i>	32.09	29.96	25.15	0.046
<i>Had a fantasy of doing something terrible (e.g., torturing) to someone?</i>	59.6	41.32	34.81	<0.0005***
<i>Fantasized about having violent non consensual sex with someone?</i>	35.05	31.49	27.25	0.029
<i>While an adult, had sexual desires for a minor?</i>	28.63	28.35	22.89	0.068
<i>Neglected to tell a partner about a sexually transmitted disease from which you were currently suffering?</i>	4.37	3.88	1.8	0.057
<i>Had sex with someone who was too drunk to know what they were doing?</i>	11.86	8.41	5.7	0.002*
<i>Stolen anything worth more than \$100?</i>	11	10.65	6.88	0.046
<i>Tried to gain access to someone else's (e.g., a partner, friend, or colleague's) email account?</i>	30.41	33.85	23.38	0.001**
<i>Looked at pornographic material?</i>	92.65	90.74	89.02	0.139
<i>Made a false insurance claim?</i>	4.89	5.34	2.55	0.061
<i>Cheated on your tax return?</i>	18.29	19	21.1	0.504
<i>Claimed to have education that you didn't actually have?</i>	6.75	9.91	6.63	0.081
<i>While in a relationship, flirted with somebody other than your partner?</i>	74.23	75.98	65.37	<0.0005***
<i>Taken credit for someone else's work?</i>	16.16	19.42	12.06	0.005
<i>Known about or witnessed a serious crime and failed to report it or stop it?</i>	7.79	7.75	5.26	0.192
<i>Let a friend drive after you thought he or she had had too much to drink?</i>	48.97	54.44	37.48	<0.0005***
<i>Made up a serious excuse, such as grave</i>	35.66	30.2	21.98	<0.0005***

<i>illness or death in the family, to get out of doing something?</i>				
<i>Lied about your income to someone?</i>	34.78	38.26	31.03	0.051
<i>Called in sick when you were not sick?</i>	70.81	68.99	60.8	0.001**
<i>Visited an internet dating website, even just to check out what types of people might be available?</i>	50.83	53.97	45.4	0.021
<i>Pretended not to see a beggar to avoid being seen as stingy?</i>	74.12	79.17	65.58	<0.0005***
<i>Downloaded pirated songs from the Internet?</i>	61	59.25	53.92	0.058
<i>Gone on a date only to make somebody else jealous?</i>	12.06	15.56	10.94	0.069
<i>Drunk so much that you got a hangover?</i>	83.82	81.85	75.37	0.002*
<i>Littered in a public space?</i>	64.02	65.67	49.72	<0.0005***
<i>Failed to do chores in a shared house or apartment?</i>	75.37	78.25	71.67	0.047
<i>Failed to tip a waiter in a country in which tipping is customary?</i>	34.38	32.14	33.58	0.751
<i>Failed to turn the lights out at home or work, just because you were feeling lazy?</i>	78.15	79	79.63	0.848
<i>In the last year, eaten meat, poultry, or fish?</i>	94.98	96.83	97.2	0.135

Notes: Questions presented in decreasing order of intrusiveness. Percents are of participants who provided an answer. Last column presents 3-way Pearson $\chi^2(2)$ *p*-values, including Bonferroni correction for *n* = 30; *: *p* < .10; **: *p* < .05; ***: *p* < .01.

Table 6 – Study 2B: Random effects ordered probit estimates

	Study 2B	
	DV: Intrusiveness	
	<i>Coefficient</i>	<i>p</i>
Constant	.2044809	0.000
<i>Decreasing Conditions</i>		
Increasing Conditions	.481771	0.000
<i>Tame Questions</i>		
Intrusive Questions	1.258393	0.000
	<i>Prob > χ^2 =</i> <i>0.0000</i>	
	<i>n=123</i>	

Table 7 – Studies 2C and 2D: Random effects probit estimates

	Study 2C DV: Engaged in Behavior		Study 2C Reduced conditions set DV: Engaged in Behavior		Study 2D Wordiness Conditions DV: Engaged in Behavior		Study 2D Intrusiveness Conditions DV: Engaged in Behavior	
	<i>Coefficient</i>	<i>p</i>	<i>Coefficient</i>	<i>p</i>	<i>Coefficient</i>	<i>p</i>	<i>Coefficient</i>	<i>p</i>
Constant	-0.0047733	0.928	-0.1579762	0.001	.6439973	0.000	.5868037	0.000
<i>Decreasing Condition</i>								
Increasing Condition	-0.1692851	0.019	-0.0312338	0.653	-0.2336612	0.018	-0.1151977	0.213
<i>Tame Questions</i>								
Intrusive Questions	-0.5681603	0.000	-0.5210749	0.000	-1.229157	0.000	-1.163632	0.000
<i>No Phishing Cue</i>								
Phishing Cue	-0.138234	0.038						
Phishing Cue* Increasing Condition	.0437733	0.671						
	<i>Prob > χ^2 =</i> <i>0.0000</i>		<i>Prob > χ^2 =</i> <i>0.0000</i>		<i>Prob > χ^2 =</i> <i>0.0000</i>		<i>Prob > χ^2 =</i> <i>0.0000</i>	
	<i>n=161</i>		<i>n=94</i>		<i>n=129</i>		<i>n=139</i>	

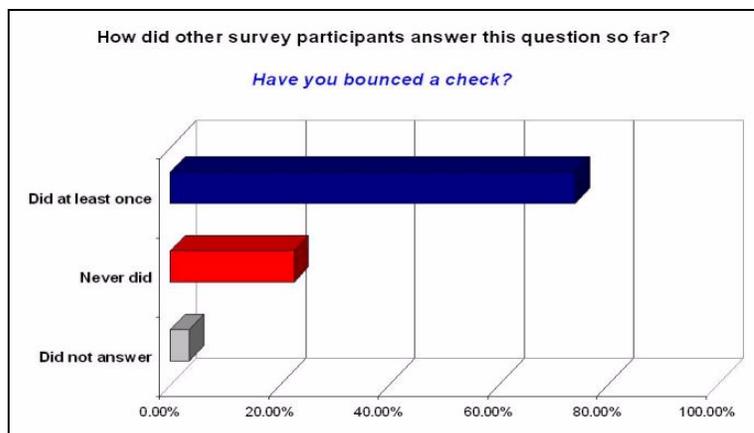


Figure 1 –Study 1A: Screenshot from the High condition

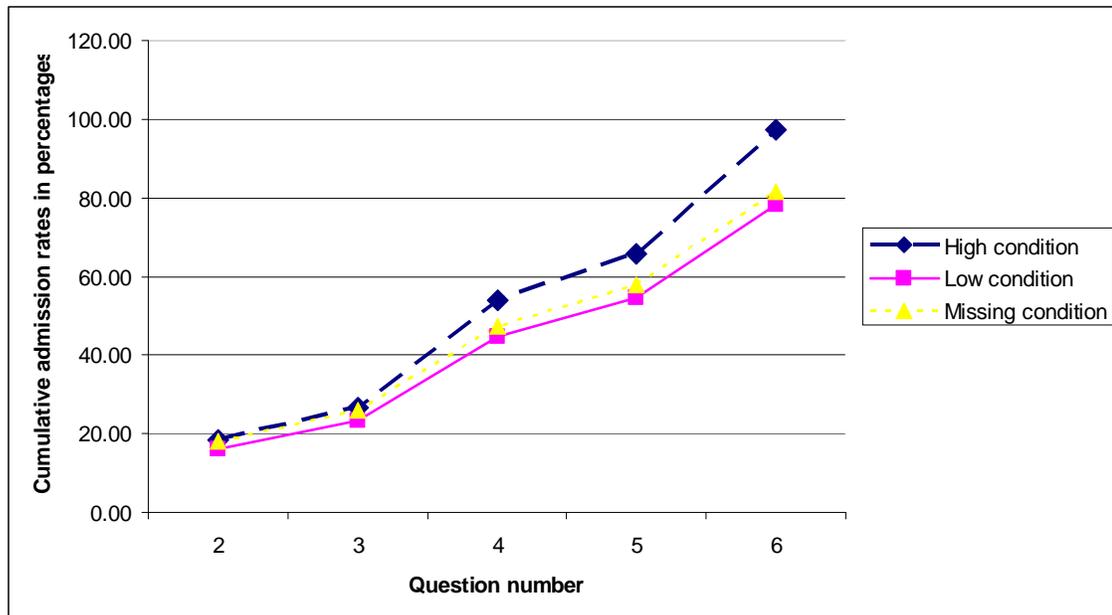


Figure 2 – Study 1A: Cumulative admission rates (in percentages) through questions 2 to 6, across conditions (the question numbers reflect the order in which the question was presented to the participants)

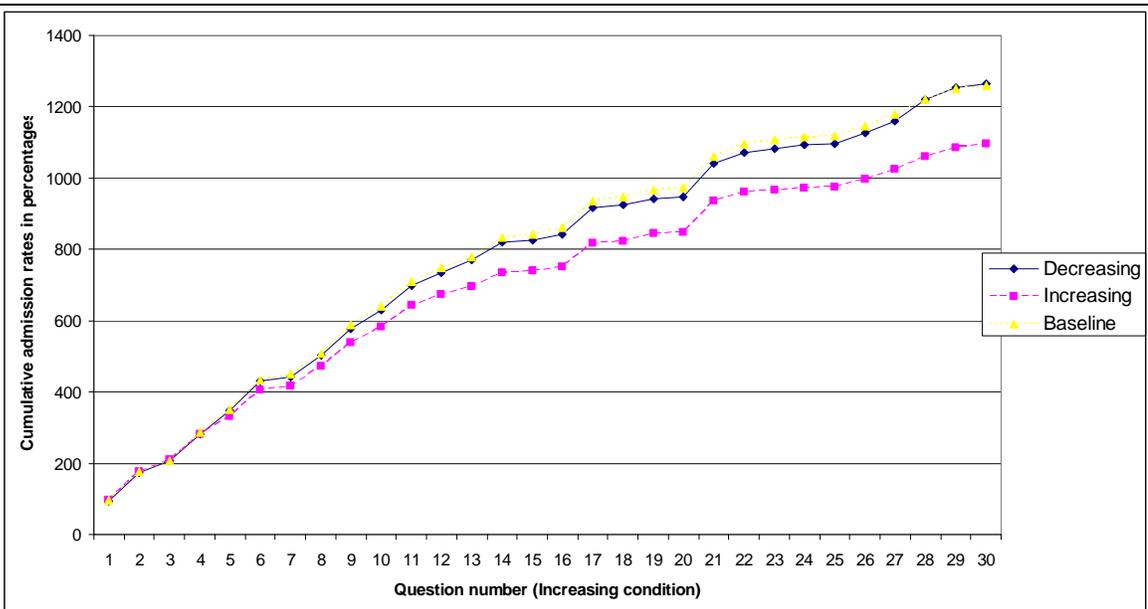


Figure 3a – Study 2A: Cumulative admission rates (in percentages), across conditions (questions are presented in order of increasing intrusiveness – as presented to participants in the Increasing conditions)

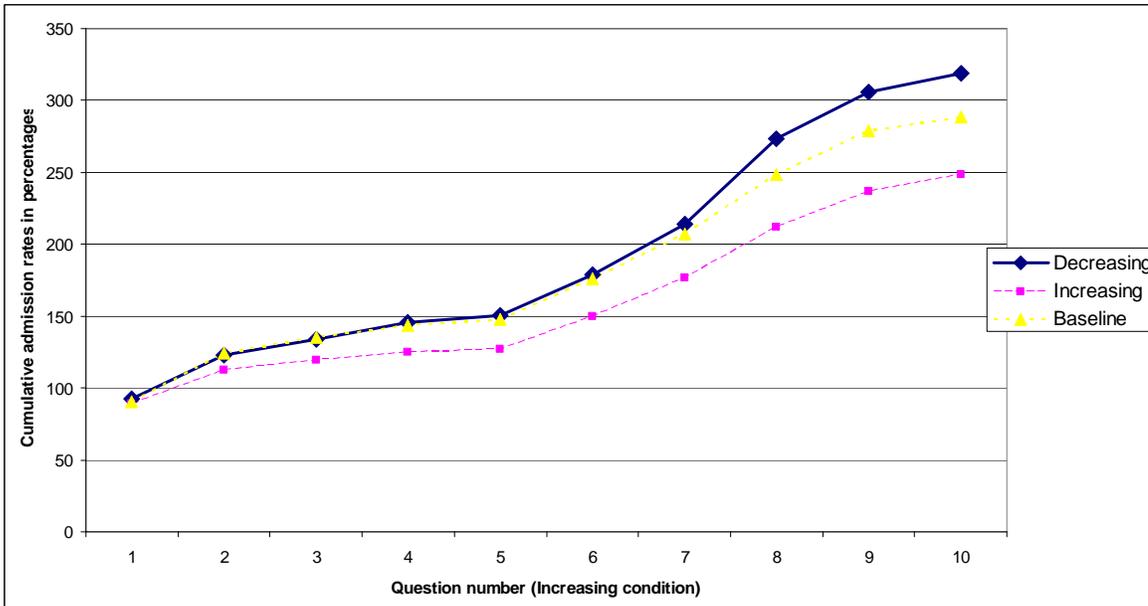


Figure 3b – Study 2A: Cumulative admission rates (in percentages), across conditions (intrusive questions only; questions are presented in order of increasing intrusiveness – as presented to participants in the Increasing conditions)

Appendix

Appendix A: Survey questions asked in Study 1A

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

Appendix B: Survey questions asked in Study 2A (as presented to participants in the Decreasing conditions)

1. *Have you had sex with the current husband, wife, or partner of a friend?*
2. *Have you masturbated at work or in a public rest room?*
3. *Have you had a fantasy of doing something terrible (e.g., torturing) to someone?*
4. *Have you fantasized about having violent non consensual sex with someone?*
5. *While an adult, have you had sexual desires for a minor?*
6. *Have you neglected to tell a partner about a sexually transmitted disease from which you were currently suffering?*
7. *Have you had sex with someone who was too drunk to know what they were doing?*
8. *Have you stolen anything worth more than \$100?*
9. *Have you tried to gain access to someone else's (e.g., a partner, friend, or colleague's) email account?*
10. *Have you looked at pornographic material?*
11. *Have you made a false insurance claim?*
12. *Have you cheated on your tax return?*
13. *Have you claimed to have education that you didn't actually have?*
14. *While in a relationship, have you flirted with somebody other than your partner?*
15. *Have you taken credit for someone else's work?*
16. *Have you known about or witnessed a serious crime and failed to report it or stop it?*
17. *Have you let a friend drive after you thought he or she had had too much to drink?*
18. *Have you made up a serious excuse, such as grave illness or death in the family, to get out of doing something?*
19. *Have you lied about your income to someone?*
20. *Have you called in sick when you were not sick?*
21. *Have you visited an internet dating website, even just to check out what types of people might be available?*
22. *Have you pretended not to see a beggar to avoid being seen as stingy?*
23. *Have you downloaded pirated songs from the Internet?*
24. *Have you gone on a date only to make somebody else jealous?*
25. *Have you drunk so much that you got a hangover?*
26. *Have you littered in a public space?*
27. *Have you failed to do chores in a shared house or apartment?*
28. *Have you failed to tip a waiter in a country in which tipping is customary?*
29. *Have you failed to turn the lights out at home or work, just because you were feeling*

lazy?

30. In the last year, have you eaten meat, poultry, or fish?

Web Appendix

Web Appendix WA1: Panel data estimation approach

As noted in the main body of the paper, in order to formally test our hypotheses we estimated random effects probit and ordered probit models. This methodology allows us to estimate the effect of the treatment (such as the ostensible distribution of answers by other subjects, or the order of intrusive questions) on the probability of a question being answered admitting to the behavior (probit specifications) or on the actual reported frequency of engagement (ordered probit), taking into account the fact that the answers provided by one subject are not independent of each other, and that other subject's traits – such as her privacy sensitivity, preferences towards self-disclosure, and actual underlying rates of engagement in certain behaviors – are unobservable. In essence, we treat our data as a panel in which the unit of observation is the subject and each of his/her answers constitutes a data point. We allow for answers by the same subject to be correlated when we estimate the variance-covariance matrix of the coefficients. We assume constant correlation between any two answers within a subject (exchangeable correlation structure: Liang and Zeger, 1986).

Our model can thus be represented by the following stylized equation:

$$q_{ij} = \beta_0 + \beta_1 Treatment_i + \beta_2 Intrusive_j + \beta_3 Treatment * Intrusive_{ij} + \beta_4 Demographics_i + v_{ij}$$

$$v_{ij} = \alpha_i + u_{ij}$$

where i indexes the subject, and j indexes the question. The equation is stylized in the sense that it takes slightly different forms depending on the study presented in the main body of the manuscript. In general, *Treatment* is a dummy variable (or a set of dummy variables corresponding to a categorical variable) denoting the experimental condition(s). *Intrusive* is a dummy variable (or a set of dummy variables corresponding to a categorical variable) denoting a question's intrusiveness level (as measured in the pilot study described in the text). In some specifications, *Treatment*Intrusive* represents the interaction between the treatment and the intrusive questions. Other variables represent demographic traits.

In probit specifications, our dependent variable q is a dummy set to 1 if a given question was answered in the affirmative (that is, the participant admitted to having engaged in the behavior at least once) and zero otherwise. While we cannot observe the true underlying frequency of engagement, or the extent participants react to the questions asked of them (we can call this the unobserved continuous variable q^*), we observe the indicator variable: whether or

not participants are willing to answer the questions in the affirmative (1) or not (0). Therefore, assuming that the error term in the equation for the unobserved variable is normally distributed, we specify a standard probit model. In *ordered* probit specifications, q is an ordinal variable representing, for instance, the self-reported frequency of engagement in a behavior (*Never, Once or twice, Sometimes, Frequently*).

Naturally, the coefficients estimated with this model do not represent the marginal effects of the explanatory variables on the probability of the question being answered in the affirmative, but they are proportional to them (the sign of the estimated coefficient will be the same as the marginal effect). In order to obtain the magnitude of the marginal effect, we can evaluate the standard normal cumulative distribution function at the estimated coefficients, and adjust for the correlation coefficient (Arulampalam 1998).

Wiji Arulampalam (1998). "A Note On Estimated Coefficients In Random Effects Probit Models," *The Warwick Economics Research Paper Series (TWERPS)*, 520).

Kung-Yee Liang and Scott Zeger (1986). "Longitudinal Data Analysis Using Generalized Linear Models", *Biometrika*, 73:13–22.

Web Appendix WA2: Demographic details for Study 1A and 2A

We present additional demographic details for Study 1A and 2A, together with some analysis of gender differences in the reaction to our experimental manipulations (Sheehan 1999, 2002 highlighted age and gender differences in privacy concerns). However, we note that all main results presented in the main body of the text for Studies 1A-C and 2A-D (which did not include demographic variables) are robust to the inclusion of demographic traits among the regressors. (The results are available from the authors on request.)

Study 1A

A total of 1,722 New York Times website visitors took the survey. Ages ranged from 18 to 81 years old (mean: 39), and participants' age distribution did not vary significantly across conditions. Gender was also similarly distributed across conditions (41% males). However, males were slightly more represented in the High condition (the difference of the gender distribution between the High and the Missing conditions is significant at the 5% level). The results presented in the main body of the manuscript do not change when we control for gender, nor other demographic variables. Race and education were also similarly distributed across conditions, with Caucasian participants, born and residing in the US representing the overwhelming majority in our sample.

We recorded participants' IP addresses, 97% of which were unique, suggesting that the participants were in fact unique visitors to the survey. (Note that non-unique IP addresses do not necessarily imply repeated users, since more than one person can use a specific computer; in any case, our results do not change when excluding the data arising from duplicate IP addresses.)

The direction of the effects reported in the manuscript is the same when controlling for gender: both males and females tend to admit more frequently in the High condition than the other conditions. However, interesting gender differences arise for specific questions. For instance, admissions in response to the question "While an adult, have you ever had sexual desires for a minor" are significantly different across conditions for males but not for females: 49.8% of males responded affirmatively in the High condition, compared to only 39.4% and 36.5% in the Low and Missing conditions, respectively (Pearson $\chi^2(4) = 19.67$, $p = 0.001$); for females, the corresponding percentages are 12.7% (High), 10.0% (Low), and 14.67 (Missing) (Pearson $\chi^2(4) = 3.04$, $p = 0.55$). The pattern is reversed in response to the question "Have you ever fantasized about having violent, non-consensual sex with someone": 35.48% of females responded affirmatively in the High condition compared to only 22.5% and 24.3% in the Low and Missing conditions, respectively (Pearson $\chi^2(4) = 13.62$, $p = 0.009$); for males, the admission

rates are similar across conditions (High: 35.9%; Low: 30.1%; Missing: 28.3%; Pearson $\chi^2(4) = 6.93$, $p = 0.14$).

Study 2A

A total of 2,310 readers of the online edition of the New York Times took part in Study 2A. Of all participants, 586 took the survey in the Decreasing conditions, 560 in the Increasing, 577 in the Random, and 587 in the Sudden. We found no significant differences in gender or age distribution across conditions. Male represented 65% of the sample; ages ranged from 18 years old to 91 year old, with a mean age of 39. Race and education were also similarly distributed across conditions. Again, Caucasian participants, born and residing in the US, represented the overwhelming majority in our sample.

Ninety-seven percent of the IP addresses of the respondents were unique, suggesting that the participants were unique visitors to the survey. Analysis of participants' IP addresses also shows virtually no overlap between the two studies: 99.6% of IP addresses were unique to each study.

As in Study 1A, we found interesting significant gender-based differences between experimental treatments in answers to specific questions. For instance, collapsing across conditions, 29.0% of males and 34.4% of females admitted to "Hav[ing] tried to gain access to someone else's (e.g., a partner, friend, or colleague's) email account." However, the difference across conditions is not significant for males, but significant for females – with, once again, (female) participants in the Random and Decreasing conditions much more likely than those in the Increasing condition to admit to the behavior (41.27% and 36.88% versus 25.00% respectively).

Kim Bartel Sheehan (1999). "An Investigation of Gender Differences in On-Line Privacy Concerns and Resultant Behaviors," *Journal of Interactive Marketing* 13(4), 24-38.

Kim Bartel Sheehan (2002). "Toward a typology of Internet users and online privacy concerns," *The Information Society*, 18(1), 21-32

Web Appendix WA3: *Treating missing answers as non-admissions*

As noted in the main body of the manuscript, missing answers are of particular importance in our Studies. They could signal that the participant dropped out of the survey, simply ignored the question(s), or, in fact, refused to answer a question – for instance, because he or she found the question offensive, or because he or she did not want to reveal personal information related to the question. Attrition could potentially generate survivor bias. On the other hand, refusal to answer a question implies the absence of an *explicit* affirmative admission. Recall that we were interested in how our manipulation would affect the subjects' propensity to explicitly admit to having engaged in sensitive behaviors. Hence, willing refusal to answer a question implies the absence of an *explicit* affirmative admission; accordingly, in some of the studies presented below we treat missing answers as non-admissions. Below, we show why the results presented in the main body of the manuscript are not determined by survivor bias and are robust to the consideration of missing answers as non-admissions.

Study 1A

In Study 1A, there was no significant difference in the (very small) percentage of participants' who abandoned the survey (Pearson $\chi^2(2) = 1.3745$, $p = 0.503$). Therefore, survivor bias does not affect our results.

In order to analyze the potential impact of missing answers on our results, we constructed a new dichotomous dependent variable, in which we combined the answer “never [engaged in the behavior]” and the absence of an answer to the question about that behavior, since both represent the absence of an explicit admission. The results presented in the main body of the text are robust to treating missing answers as non-admissions (see Table WA3.1 below). In the model without interaction, the dummies for the Low and Missing conditions are, respectively, negative but insignificant and only significant at the 10%. However, when controlling for demographic traits (recall that, in Study 1A, males were over-represented in the High condition), the dummies for the Low and Missing conditions are again significant at the 1% and 5% levels respectively (the complete results of the regression with demographic IVs are available from the authors). In the model with interaction, once again the interaction between questions and conditions are negative, and tend to have larger coefficients and more statistical power in the latter questions of the survey. Averaging over questions, participants in the High condition were still 19% more likely to admit to having engaged in the behaviors than participants in the Low condition, and also 19% more likely than participants in the Missing condition (both pairwise t -tests: $p < 0.005$).

Once again, we found no statistically significant differences between the admission rates in the Low and the Missing conditions also when treating missing data as non-admissions.

Table WA3.1 – Study 1A: Random effects probit estimates of the propensity to admit to having engaged in various behaviors, treating missing observations as non-admissions

	I		II	
	<i>Coefficient</i>	<i>P</i>	<i>Coefficient</i>	<i>P</i>
Constant	-.0752126	0.088	-.2322352	0.000
<i>High Condition</i>				
Low condition	-.0730169	0.135	.221883	0.008
Missing condition	-.0865632	0.074	.0886504	0.289
<i>Question 1 (Bouncing check)</i>				
Question 2 (Cheating on tax return)	-.9302163	0.000	-.7927455	0.000
Question 3 (False insurance claim)	-1.461808	0.000	-1.327827	0.000
Question 4 (Desire for minor)	-.6937158	0.000	-.4558649	0.000
Question 5 (Cheating on partner)	-1.272677	0.000	-1.113616	0.000
Question 6 (Fantasizing about non consensual sex)	-.5901092	0.000	-.3107695	0.000
<i>Interaction terms</i>				
Low*Q2			-.3195976	0.009
Low*Q3			-.2926674	0.039
Low*Q4			-.4216924	0.000
Low*Q5			-.322165	0.015
Low*Q6			-.4893286	0.000
Missing*Q2			-.0988805	0.411
Missing*Q3			-.1144308	0.411
Missing*Q4			-.3005261	0.009

Missing*Q5			-.1605023	0.220
Missing*Q6			-.3614744	0.002
	<i>Prob > χ^2</i>	=	<i>Prob > χ^2</i>	
	<i>0.0000</i>		= <i>0.0000</i>	
	<i>n = 1,722</i>		<i>n = 1,722</i>	

Study 2A

In Study 2A, different reasons why participants may skip certain questions interact with the order in which participants faced those questions across different conditions. For instance, participants may not provide an answer to a question because they refused to address it, due to its intrusiveness; or, they may have simply dropped out of the survey (Reips 2002). Since participants in Study 2A (unlike those in Study 1A) may therefore end up skipping questions of *different* sensitivity across the different conditions (thereby confounding the admission rates by condition), we need to take the position of the question within the survey into explicit account in our analysis in order for control for possible survivor bias as well as to interpret the meaning of missing answers.

Survivor bias. First of all, we contrasted, across conditions, the share of participants who did not answer a question based on the question's *position* in the questionnaire, rather than its intrusiveness level: 13.65% of participants in the Decreasing condition skipped their very first question (with most of them thereafter skipping the remaining questions as well), versus just 4.29% in the Increasing condition (with most of them also skipping their remaining questions): Pearson $\chi^2(1) = 30.44$, $p < 0.0005$. However, after the initial gap between Decreasing and Increasing conditions, the percentages of participants who skipped questions follow virtually identical patterns in both conditions, with their numbers gradually spiking in correspondence of html page changes (that is, every sixth question) as the surveys progress. By the end of the questionnaire, 18.43% of participants in the Decreasing condition skipped the very last question, and 10.89% of those in the Increasing condition did so (Pearson $\chi^2(1) = 12.94$, $p < 0.0005$). Importantly, the difference in the gaps between proportions of skipped questions across Increasing and Decreasing conditions at the start of the survey (9.36%) and at the end (7.54%) is *not* statistically significant.

Consider Figure WA3.1, where questions are represented on the x-axis in the order in which participants across different conditions saw them. Note that after the initial differences,

the percentage of participants who skip questions increases as the survey goes on, following virtually identical patterns for the Decreasing and Increasing conditions (the jumps correspond to the points where participants had to switch to the next html page: questions 7, 13, 19, and 25). The difference between the percentages of participants declining to respond to questions (yellow line) slightly decreases along the questionnaire – as participants in the Increasing conditions start facing more and more intrusive questions, while participants in the Decreasing condition start facing tamer questions.

This set of results tells us two important things. First, a relatively larger number of participants in the Decreasing condition were put off by the intrusive questions asked at the very beginning of their survey. (Still, *even when we include those subjects in the analysis*, by treating their missing observations as non-admissions, the Decreasing condition still exhibits higher admission rates for the intrusive questions than all other conditions; this means that survivor bias is not determining the results we presented in the main body of the text; we further account for this below in this section.) Second, after such initial difference, the propensity of participants in the Decreasing condition to provide an answer to their *subsequent* questions was identical to that of participants in the Increasing condition, suggesting that the subsequent observed patterns of missing answers are closer to known “answering drop-outs” dynamics in online surveys (Bosnjak and Tuten 2001), than differential patterns of self-selection across conditions.

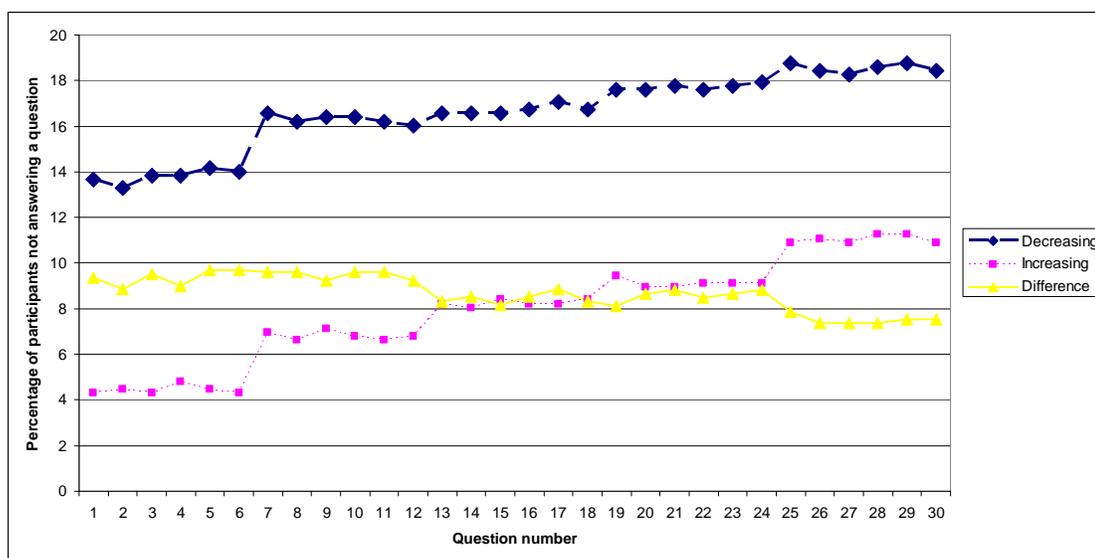


Figure WA3.1 – Study 2A: Proportion of skipped answers, by condition (questions are numbered in the order in which participants across different conditions saw them)

Treating missing answers as non-admissions. We controlled for variations across conditions in the patterns of skipped questions also by treating, as noted in the manuscript, the lack of an

answer as an absence of admission. We found that the aggregate admission rates between conditions remain statistically significant (see Table WA3.2 below). Participants in the Increasing condition are still 15% less likely than those in the Random condition to admit to having engaged in the various behaviors (t-test on mean admission rates: $p = 0.0006$). However, the differences in overall admission rates between the Increasing and Decreasing conditions fade (null hypothesis that the mean overall admission rates are not the same across the two conditions: $p = 0.34$). If anything, in terms of overall admission rates, participants in the Decreasing are marginally (4%) *less* likely to admit to the various behaviors than those in the Random condition ($p = 0.027$). The reason is simple: Recall that, in the Decreasing Condition, a higher number of participants were turned off by the initial questions and left the survey. Hence, those participants also did not answer the remaining questions throughout the survey. This, in turn, inflates the number of missing observations for the same questions (which, in the Decreasing condition, arrived *at the end* of the questionnaire).

However, and importantly, admission rates to *intrusive* questions remain statistically significantly lower in the Increasing condition than in all others conditions, and significantly higher in the Decreasing condition than in all others (see the second set of columns in Table WA3.2A, which presents the results of a variant of the random effects probit model in which we treat missing observations as non-admissions but also add the interaction terms; the results confirm the significant interaction between conditions and question intrusiveness, and the higher propensity of participants in the Decreasing condition to admit to sensitive behaviors). As noted above in this same section, even though more subjects dropped from the survey from the start in the Decreasing condition than in the other conditions, when we include them in the analysis by treating their missing observations as non-admissions, participants in the Decreasing condition are still 43% more likely to admit to the sensitive behaviors than those in the Increasing conditions, and 10% more likely than those in the Random conditions (t-tests on admission rates: $p < 0.0005$ and $p = 0.0643$ respectively). Furthermore, even when we (conservatively) compare admission rates in the Decreasing condition treating their missing answers as *non-admissions*, to admission rates in the Increasing condition *without* treating their missing answers as non-admissions, we still find that participants in the former conditions are 29% more likely to admit to the sensitive behaviors.

Table WA3.2 – Study 2A: Random effects probit estimates of the propensity to admit to having engaged in various behaviors, treating missing observations as non-admissions

	I		II	
	<i>Coefficient</i>	<i>p</i>	<i>Coefficient</i>	<i>p</i>
Constant	.0643993	0.020	-.0006412	0.983
<i>Decreasing condition</i>				
Increasing condition	.0181478	0.626	.1339463	0.001
Random condition	.0893152	0.020	.1685371	0.000
<i>Tame</i>				
Moderate	-.833524	0.000	-.7691731	0.000
Intrusive	-.885291	0.000	-.7398104	0.000
<i>Interaction terms</i>				
Increasing*Moderate			-.1360508	0.000
Increasing*Intrusive			-.2369674	0.000
Random*Moderate			-.0561135	0.123
Random*Intrusive			-.1998099	0.000
	<i>Prob > χ^2 = 0.0000</i>		<i>Prob > χ^2 = 0.0000</i>	
	<i>n=1,723</i>		<i>n=1,723</i>	

Michael Bosnjak and Tracy L. Tuten (2001), “Classifying Response Behaviors In Web-Based Surveys,” *Journal Of Computer-Mediated Communication*, 6(3).

Ulf-Dietrich Reips (2002). “Standards For Internet-Based Experimenting,” *Experimental Psychology*, 49(4), 243-256.

Web Appendix WA4: *Email provision*

Study 1A

Participants across all conditions had the opportunity to provide their email addresses before questions about behaviors were asked, but after a page reminding them about the sensitivity of their answers. The vast majority of participants provided email addresses (65.45 %, 71.99%, and 68.89% in the High, Low, and Missing conditions respectively, with no significant difference across conditions). Although the percent giving an email address was, by chance, lowest in the High condition, all the results presented in the main body of the manuscript remain significant if we control for the provision (or lack thereof) of email addresses in the regressions (see Table WA4.1 below), or if limit analyses to the subset of participants who provided their email addresses at the beginning of the study; indeed, in certain cases, the statistical significance of the tests increases. Choosing not to provide an email address can be interpreted as a possible sign of higher privacy sensitivity: these participants were also generally less likely to admit to sensitive behaviors than participants who provided their email addresses, regardless of the condition (t-test on mean admission rates: $p < 0.05$). Accordingly, when we limit the analysis to the subset of participants who did *not* provide email addresses, the differences between conditions in propensity to admit decrease – privacy sensitive participants (who chose not to disclose identifying information) were *less* affected by the manipulations. Overall, therefore, participants who provided email addresses were not only more susceptible to the manipulation, but were also more likely to admit to having engaged in the behaviors (collapsing across conditions).

Table WA4.1 – Study 1A: Random effects probit estimates of the propensity to admit to having engaged in various behaviors, including email provision as IV

	I		II	
	<i>Coefficient</i>	<i>P</i>	<i>Coefficient</i>	<i>p</i>
Constant	.0017701	0.03	-.153407	0.026
<i>High Condition</i>				
Low condition	-.1132629	0.013	.1844988	0.029
Missing condition	-.1000169	0.028	.0673135	0.424
<i>Question 1 (Bouncing check)</i>				

Question 2 (Cheating on tax return)	-0.94517	0.000	-0.8129744	0.000
Question 3 (False insurance claim)	-1.468798	0.000	-1.345174	0.000
Question 4 (Desire for minor)	-0.6821115	0.000	-0.4392177	0.000
Question 5 (Cheating on partner)	-1.266292	0.000	-1.11512	0.000
Question 6 (Fantasizing about non consensual sex)	-0.5604563	0.000	-0.2805655	0.001
<i>Email not provided</i>				
Email provided	.1266368	0.005	.1268141	0.005
<i>Interaction terms</i>				
Low*Q2			-0.3160481	0.011
Low*Q3			-0.2826862	0.048
Low*Q4			-0.4387921	0.000
Low*Q5			-0.3103734	0.021
Low*Q6			-0.4945018	0.000
Missing*Q2			-0.0832976	0.497
Missing*Q3			-0.0898758	0.523
Missing*Q4			-0.145172	0.014
Missing*Q5			-0.3527616	0.276
Missing*Q6			-0.3527616	0.003
	<i>Prob > χ^2 =</i> <i>0.0000</i>		<i>Prob > χ^2 =</i> <i>0.0000</i>	
	<i>n = 1,538</i>		<i>n = 1,474</i>	

Study 2A

In Study 2A, we also manipulated whether email addresses (and therefore *potentially* identifying information) were asked before or after the actual survey, and crossed this manipulation with the order in which questions of different intrusiveness were presented to participants. We found no significant difference, across conditions, in the propensity to reveal an

email address *before* the questionnaire, nor, more interestingly, any significant difference across conditions in the propensity to reveal an email address *after* the questionnaire. However, collapsing across conditions, participants were more likely to provide an email address when it was asked at the beginning than when it was asked at the end of the questionnaire (81.78% vs. 65.75%; Pearson $\chi^2(1) = 56.56, p < 0.0005$).

The differences in admission patterns across the Decreasing, Increasing, and Random conditions which we described above are the same regardless of the point at which email addresses were requested (see Table WA4.2 below). However, we also detected slightly higher admission rates across all question types (as well as when considering separately the intrusive, tame, and moderate questions) for participants who *did* reveal email address over those who did not (overall admission rate for participants who provided an email, collapsing across conditions: 0.41; overall admission rate for participants who did not provide an email, collapsing across conditions: 0.38; t-test on mean admission rates: $p < 0.003$). This difference is significant *only* for conditions in which email addresses were asked at the end ($p < 0.0006$). In other words, participants with higher admission rates also tended to be more likely to provide an email address at the end. This suggests that there may be systematic differences between individuals in their concern for privacy, with those less concerned about privacy more likely to both provide their email address and admit to having engaged in sensitive behaviors.

Table WA4.2 – Study 2A: Random effects probit estimates of the propensity to admit to having engaged in various behaviors, including email provision as IV

	I		II	
	<i>Coefficient</i>	<i>p</i>	<i>Coefficient</i>	<i>p</i>
Constant	.3242302	0.000	.2923509	0.000
<i>Decreasing condition</i>				
Increasing condition	-.1566914	0.000	-.1136182	0.000
Random condition	-.0129186	0.585	.0342294	0.283
<i>Tame</i>				
Moderate	-.8510236	0.000	-.8371954	0.000
Intrusive	-.9093745	0.000	-.8309501	0.000
<i>No email provided</i>				
Email provided	.0579811	0.013	.0589626	0.012
<i>Interaction terms</i>				
Increasing*Moderate			-.0357236	0.339

Increasing*Intrusive			-.095037	0.011
Random*Moderate			-.0044336	0.906
Random*Intrusive			-.1375842	0.000
	<i>Prob > χ^2 = 0.0000</i>		<i>Prob > χ^2 = 0.0000</i>	
	<i>n=1,581</i>		<i>n=1,425</i>	

Web Appendix WA5: *Impact of admitting behaviors in Study 2A*

As we noted in the main body of the manuscript, participants in the Decreasing condition admitted to sensitive behaviors more frequently than participant in the Increasing condition – who started the questionnaire with much tamer question, and may have adapted their expectations about the sensitivity of the questionnaire to those questions. An issue worth exploring is whether the disclosure behavior of a subject at the beginning of the survey ends up affecting, endogenously, his or her propensity to admit to behaviors in the rest of the survey, and whether this impact is different across conditions. In other words: on top of the impact of the order of questions, does the way the initial questions in a survey are *answered* impact how following questions will be answered?

Naturally, we would expect that to be the case – since propensity to disclose may be a subject’s trait that influences the subject’s behavior across the entire survey. The issue, however, is whether subjects in different conditions reacted differently in this respect. To address this issue, we regressed the mean admission rate to the four most intrusive questions for the participants in the Increasing conditions (therefore, their last four in the questionnaire) over their admission rate to the four tamest (their first four in the questionnaire). The coefficient was positive (0.1297) and significant at $p = 0.013$. We then regressed the mean admission rate to the four tamest questions for the participants in the Decreasing conditions (therefore, their last four in the questionnaire) over their admission rate to the four most intrusive (that is, their first four in the questionnaire). The coefficient was also significant at $p = 0.000$ and of similar magnitude (0.1599). This result suggests that, in both cases, the propensity to admit to questions at the beginning of the survey is positively correlated with the propensity to admit to behavior at the end of the survey -- which is plausible. This correlation holds true both for subjects in the Decreasing condition and for those in the Increasing condition.