Self-disclosures, impression formation, and biases in Web 2.0

A dissertation submitted to the
Heinz College – School of Public Policy and Management
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy in Public Policy and Management

by

Laura Brandimarte

Dissertation Committee:

Alessandro Acquisti (Chair)
George Loewenstein
Francesca Gino

Carnegie Mellon University
Pittsburgh, Pennsylvania

December, 2012

Self-disclosures, impression formation, and biases in Web 2.0

Abstract

The first decade of the 21st century has witnessed the rise of Web 2.0 technologies that allow users to create and share content online with friends – and strangers alike. These technologies have generated an 'enthusiasm for sharing' as well as privacy concerns that researchers, organizations and policy makers often measure and debate. After introducing the trade-offs of information sharing (Chapter 1), I investigate one of the antecedents of willingness to share personal information – namely, perceived control – as a possible explanation for the success of online social media, notwithstanding people's diffuse concern for privacy protection. The results challenge the consensus view of control as a sufficient means of privacy protection, since increased perceived control may lead to higher self-disclosure, even in situations of higher objective risks (Chapter 2). I then examine some of the consequences that public disclosures entail in terms of impression formation and reputation building for the person who discloses information. It could be argued that online disclosures, even if embarrassing or incriminating, will soon stop representing diagnostic information about people – either because we will forget about them as time passes (Chapter 3), or because, as social norms about disclosure change, the majority of people will soon be tainted by embarrassing online records, which may contribute to render them more lenient towards others' disclosures (Chapter 4). The results of studies in Chapters 3 and 4 challenge this view, and suggest that online disclosures may have a long lasting effect on impression formation.

Acknowledgments

This work and the completion of my PhD program would not have been possible without the endless support of my adviser, Prof. Alessandro Acquisti. To him I owe most of the things I know now and most of the research skills I have acquired through the years. He went way beyond what is expected of an adviser, always pushing and motivating me, standing by my side unconditionally. Thank you, AA, for these fantastic years of work, travels and fun! I'll be looking forward to many more challenges and collaborations!

Lots of thanks to the other members of my *stellar* dissertation committee, as some have rightly defined it, Prof. George Loewenstein and Prof. Francesca Gino, who have provided inspiration, invaluable research advise, personal support and warmth. You are my models.

Special thanks to my family, who have always believed in me, and made many sacrifices just so I could follow my dream, which I'm sure was also Dad's. Gabry, Fily and Lale: we went through a lot, but we made it come true. My PhD and anything that will come of it in the future is dedicated to you all, I couldn't have accomplished any of this without you.

Thanks to all the friends who have been working with me these past few years, either as co-authors, collaborators or research assistants. To all the members of the fabulous PeeX Lab: you guys are truly something else. I have learned so much through our endless discussions and brainstorming sessions!

Many, many, many and more thanks to Gretchen, who has always taken care of us students, so that we could concentrate on our little research and worry about nothing else.

A heartfelt thank you and lots of love to my guardian angels, Camillo, Claudio and Rosa.

And thanks to all my friends and relatives, who would be too many to fit in one page, for putting up with this student-forever, and for bringing uncountably many moments of joy, laughter and fun – especially when AS Roma, Federer or Rossi won. Big hug to my favorite boys, the Testa kids, and to my "little" cousin Regina: be strong and be brave, you all! To the Soverino's, the Perfetti's and the Galassetti's: your friendship is a gift and an honor, you are family. La Sapienza group, who have been rooting for me since I left home, deserve a special mention, together with my mentor from the days of Istituzioni di Economia Politica: Gustavo, thanks for being so passionate about all things Econ, for transmitting to me the love and dedication to teaching, and for believing in me since that first infamous 27/30!

Grazie ragazzi! Vi voglio bene!

#TeamLaura

Table of Contents

Chapter 1

Trade-offs of information sharing technologies

The Internet has brought about a revolution in the way people communicate, work, interact, learn, and entertain themselves (Hilbert & Lopez, 2011; Litan & Rivlin, 2001; Viard & Economide, 2011). Information, with the power it grants to its possessor, is at the center of this revolution (Floridi, 2007, 2010). Information has been a prominent element in economics research since seminal articles by Hayek (1945), who discussed the asymmetries of 'knowledge' across the various actors in the economy, and Stigler (1961), who stressed the importance of accounting for 'the cold winds of ignorance' in order to understand the facts of economic choices. Information is still a crucial research topic today, in economics as much as in decision making and other social sciences (Brown & Duguid, 2000). In his address at the 1997 Global Knowledge Conference in Toronto, Secretary General of the United Nations, Mr. Kofi Annan, stressed how "knowledge is power. Information is liberating," and how information "has a great democratizing power waiting to be harnessed to our global struggle for peace and development."[1]

If information sets people free – free to think, to decide, to choose – then any technology that makes information publicly available, easy to search and at very low cost is the key to freedom – and the Internet should represent a passe-partout, opening all doors to information and knowledge. But there is a catch. On the one hand, the Internet gives people an arguably unprecedented amount of freedom of expression and communication; on the other hand, the Internet may also constrain people if it makes their private information more likely to be diffused

[1] Global Knowledge 97, Toronto, Canada, June 22, 1997. Text of the address available at http://www.deepsky.com/~madmagic/kofi.html.

in ways that may thwart future opportunities. “As people use the freedom-enhancing dimensions of the Internet, as they express themselves and engage in self-development, they may be constraining the freedom and self-development of others – and even of themselves” (Solove, 2007).

Much like other revolutionary technologies, the Internet was created to respond to certain human necessities, but then its evolution and developments ended up exacerbating those necessities and even generating new ones. One of the main motivations of the engineers and computer scientists, who created the Internet back in the late 60s and early 70s, was the need to get connected to each other and work together without being physically in the same place at the same time. Being able to access and share information from various locations increased productivity significantly. The initial connection of two or three research centers soon became a worldwide network, available for commercial use.

When Web 2.0 technologies came along, it became possible to not only passively use the Internet as a repository of information to be browsed, but also to actively generate and share content of any kind (Klamma, Cao, & Spaniol, 2007; Rafaeli & Raban, 2005) without any prior programming knowledge or technical skills. These kinds of technologies in particular, with their high level of user engagement, have profoundly modified the scenario of information sharing – especially *personal* information (e.g. Strater & Lipford, 2008; for a study focused on adolescents, see Hinduja & Patchin, 2008). What was once considered private – either unworthy of, or too worthy for public consideration – has suddenly become the prominent content of social media, where millions of users get connected. A widespread urge to constantly share, sometimes broadcast details of one’s personal life can be observed now, that didn’t seem to exist prior to the advent of social networking sites, weblogs, photo or video sharing sites. In their dialectic

interaction, society and technology influence each other: society stimulates technological innovation, but when technology gathers "momentum," it can shape and affect habits and norms much more than it is affected by them (Hughes, 1994).

Human beings are social by nature: it is only natural that they search for new, effective ways to interact with each other and share personal experiences. "The more we know about each other, the greater the chances that we will survive," said media tycoon Bill Parrish in the 1998 movie "Meet Joe Black." But information sharing is characterized by an inherent trade-off, which technologies of this hyper-connected world on the one side, and limitations and biases that human decision making is subject to on the other side, are making especially hard to recognize and, therefore, even harder to resolve. Opening up towards others may have unmatched benefits – from receiving support by similar others, to developing a sense of belonging to a group; from venting one's frustrations and disappointments, to sharing one's happiness and successes with beloved ones – but these benefits come to the necessary cost of giving up parts of one's private, intimate world, and of dynamically adjusting the boundaries of privacy (Nissenbaum, 2004; Petronio, 2002) that separate the self from the rest of the world.

Social media, and information sharing in general, have certainly met the human need for connection, but they have generated other needs, such as the need to feel in control over personal information (Phelps, Nowak, & Ferrel, 2000); the need for secure platforms, where revelations and transactions do not lead to unwanted targeted advertising, stalking, identity theft, or credit card fraud (Anton, Earp, & Young, 2010; McDonald & Cranor, 2010); the need to build and protect a certain image for one's online persona (boyd, 2004; Stutzman, 2006; Walker 2000); the need to keep that online persona as a separate entity from the real self, still maintaining one's true identity (Bargh, McKenna, & Fitzsimons, 2002; Zhao, Grasmuck, & Martin, 2008).

The shift towards more openness that has characterized the last few years of Internet interactions is not necessarily a sign of decreased privacy needs. Quite the opposite, polls and surveys (ConsumersUnion.org, 2008; Federal Trade Commission, 2000; Harris Interactive, 2002) where privacy concerns are measured typically show high levels of worry and unease, especially in the online world. This shift is thus more likely the result of the difficulties in discerning and correctly assessing the risks of online disclosures, especially if compared to their more immediate and salient benefits (Acquisti, 2004; Acquisti & Grossklags, 2005).

The privacy trade-offs associated with online disclosure of personal information are often hard to identify. First of all, surfing the web is perceived as an essentially *free* experience: once the monthly fee is paid to the Internet Service Provider, and with the exception of some paid subscription services, anybody can post information that can be easily accessed free of charge. Agreeing to the Terms of Service of a website or providing a personal email address is hardly interpreted as a form of payment for the access and use of that website. Recently, expressions have been circulating on the Internet that make the transaction implicitly completed when a website is visited more evident. Sayings such as "If you're not paying for it, you're the product," or "when you're not paying with cash you're paying with your personal information,"[2] have appeared in several blogs to clarify the misconception of free online experiences. But, like research on self-regulation shows (for an overview, see Boekaerts, Pintrich, & Zeidner, 2000), conscious awareness may not be enough to affect behavior: browsing the Internet may still be perceived as a free experience even though one knows it isn't (for a seminal paper on how feelings and emotions affect behavior and choices, at times more strongly than cognitive assessments, see Loewenstein, Weber, Hsee, & Welch, 2001; for an overview, see Slovic, 2010).

[2] See for instance: http://lifehacker.com/5697167/if-youre-not-paying-for-it-youre-the-product, last accessed on November 19, 2012.

Second, privacy policies of web services, that are supposed to explain the privacy trade-off and make it salient and transparent, in order for the reader to make an informed decision about the release of personal information, are hard to find – often hidden at the bottom of the web page in small font – and even harder to understand (McDonald & Cranor, 2008). This implies that even privacy-concerned users will have difficulties in understanding what exactly they are giving up in order to obtain the services of a particular website, and under what exact conditions.

Third, Internet users may not realize that they are paying for a service at the time they use it because costs of disclosure often come in the future, in the form of unwanted advertisements, price discrimination, spam, or worse, identity theft. Moreover, even if users were perfectly aware of such future costs, and they were 'privacy fundamentalists' (Westin & Harris Louis and Associates, 1991), they would be likely to under-weigh them as compared to the corresponding immediate benefits of the online experience, and resolve the trade-off too easily in favor of self-disclosure (Acquisti, 2004; Acquisti & Grossklags, 2004).

Finally, given the abstract and multi-faceted nature of privacy, privacy preferences are hardly stable: they are strongly influenced by context and they can be easily manipulated through framing (John, Acquisti, & Loewenstein, 2011; Acquisti, John, & Loewenstein, 2012). Thus, the combination of features of information sharing technologies, such as poor presentation of privacy policies or trustworthy 'look and feel,' and cognitive and behavioral biases, such as hyperbolic discounting and self-control issues, make privacy trade-offs quite hard to resolve in a rational and utility-maximizing fashion.

In this dissertation, we use methods and concepts of behavioral decision research to analyze the implications of new information sharing technologies for privacy protection and impression formation. In Chapter 2, we focus on a particular feature of online information sharing technologies – namely, granular privacy controls – that is meant to help users share personal information while still preserving their privacy, but that, paradoxically, could make privacy protection actually harder to achieve. We introduce a new reason, which we dub the "control paradox," that could explain the widely observed inconsistencies between stated privacy preferences and actual behaviors. We show how the control paradox may lead to a resolution of the privacy trade-off that leans too much in favor of disclosure at the expense of privacy protection.

In Chapters 3 and 4, we analyze the consequences of such resolution in terms of the impressions others will form of a person based on public disclosures. In particular, we challenge the view that online disclosures will be granted little consideration in the future, as people will quickly forget about them, or as most people will have their own embarrassing online record of information. Instead, we argue that impressions generated from online disclosures are "sticky," and may be hard to modify even as time passes and social norms change.

Specifically, in Chapter 3 we investigate the role of perceived information diagnosticity in impression formation. We find that if information is perceived as diagnostic of someone's personality, it will maintain its relevance for impression formation even if it refers to something that occurred way back in the past. With embarrassing or inappropriate disclosures becoming more common in online social media, this finding has a key implication for impression formation: embarrassing information, even though posted as a teenage joke, may constitute a

permanent taint in our past, a spot that will strongly affect what others think about us and the way in which they evaluate us.

Finally, in Chapter 4 we study the effect of evolving social norms on impression formation based on online disclosures. Public disclosure of embarrassing or sensitive information may have become more common and socially acceptable, but this does not necessarily imply that the way in which such information is combined in order to form an impression of a target person changed accordingly, or that such disclosures will stop affecting impression formation altogether. In three experiments, we study the effect of disclosure of a certain trait or behavior, simply embarrassing or even unethical, on the impression that one will form of others who made similar disclosures. We find that admitting embarrassing or unethical behaviors can paradoxically make one even more judgmental towards others with a similar history of disclosures.

Chapter 2

Misplaced confidences: Privacy and the control paradox

with Alessandro Acquisti and George Loewenstein

Abstract

We test the hypothesis that increasing individuals' perceived control over the release and access of private information – even information that allows them to be personally identified – will increase their willingness to disclose sensitive information. If their willingness to divulge increases sufficiently, such an increase in control can, paradoxically, end up leaving them more vulnerable. Our findings highlight how, if people respond in a sufficiently offsetting fashion, technologies designed to protect them can end up exacerbating the risks they face.

INTRODUCTION

A key concern in debates about privacy is whether people are able to navigate issues of sharing and protecting personal information to their own advantage. The general assumption, which we endorse, is that policy intervention is motivated to the extent that people are poor navigators. Much as seat belts in cars are justified by the fact that people's natural driving habits (as well as those of other drivers) create an unacceptable level of risk, privacy interventions can be justified by similar limitations of individuals' abilities to manage privacy-related risks. Indeed, in recent years, considerable evidence has emerged that individuals' privacy decision making is far from optimal, and is subject to various non-normative influences. For example, privacy assurances can have the perverse effect of causing people to 'clam up,' whereas cues that divulgence could be risky, such as a survey's informal feel, can cause them to reveal information exactly when the situation warrants self-protective concealment of information (John, Acquisti & Loewenstein, 2011).

The analogy to seatbelts, however, raises an important caveat, which is the central focus of the current paper. Although seatbelts certainly save lives, they don't save as many lives as would be expected based on the effectiveness of the technology itself, and, research suggests, they have increased fatalities among pedestrians and bicyclists (Semmens, 1992; Wardlaw, 2000). The reason is that people who wear seatbelts tend to drive more recklessly. More generally, people often respond to safety measures intended to protect them in ways that counteract the protection – a phenomenon known as 'risk homeostasis' or, more colloquially, the 'Peltzman Effect' (Peltzman, 1975).

In this paper, we explore an analogous phenomenon in the realm of privacy. In response to the common perception that consumers are increasingly concerned about their privacy, particularly in today's Internet age, industry organizations, policy makers, and even privacy advocates have promoted solutions that involve giving individuals more control over their personal information. Consistent with a Peltzman Effect, however, we document a 'control paradox' such that people who experience more perceived control over limited aspects of privacy sometimes respond by revealing more information, to the point where they end up more vulnerable as a result of measures ostensibly meant to protect them. On the other hand, lower perceived control can result in lower disclosure, even if the associated risks of disclosure are lower.

Prior research has identified control as a determinant of risk perception and risk taking (e.g., Harris, 1996; Klein & Kunda, 1994; Nordgren, Van der Pligt, & Van Harreveld, 2007; Slovic, 1987; Weinstein, 1984): people are more willing to take risks, and judge those risks as less severe, when they feel in control. For example, people feel safer driving than flying, and as a result substitute road for air travel, in part based on the feeling that they have more control when driving. Such feelings are, in fact, often merited; people *do* have greater control over the risks they face in driving than they do over the risks they face in flying. However, driving is much more dangerous than flying, even for those who take exceptional measures to control their driving risks, because there are sources of risk that cannot be controlled, such as the larger number of vehicles driven and the behavior of other drivers. The ability to control *some* risks, therefore, seems to, in effect, obscure people's awareness of, or attention to other risks that they cannot control.

We argue that a similar misleading feeling of control underlies many instances of problematic divulgence of information, such as the publication of embarrassing or even self-incriminating information by users of online social networks, the use of social network sites by employees to denigrate their employers, and the sharing of compromising pictures on Twitter (including the notorious case of one politician). Providing control over personal information allows one to choose how much to reveal about oneself and to whom. However, much as drivers may underestimate sources of risk that do not depend on their behavior, people who feel in control of their disclosures may underestimate the level of risk that arises from other people's access and uncontrollable usage of their disclosed information, and respond by disclosing more. On the other hand, people who feel less in control of their disclosures may overestimate those risks, and respond by disclosing less.

To investigate the relationship between control, disclosure, and privacy concerns, we conducted three survey-based experiments in which respondents were asked questions that varied in sensitivity. In these experiments, we decreased (Studies 1 and 2) or increased (Study 3) participants' control over the *release* or *accessibility* of personal information. We found that perception of control affected individuals' concern about privacy, to the point that their willingness to disclose sensitive information increased, even in cases where objective risks of disclosure increased. Vice versa, lack of perceived control raised privacy concerns and caused individuals to withhold information, even in cases where objective risks of disclosure decreased.

Prior empirical studies of privacy that address issues of control have shown that lower perceived control over personal information is associated with higher privacy concerns (Xu, 2007; Hoadley, Xu, Lee,& Rosson, 2010), and that individuals who are unconcerned about privacy often explain their lack of concern by noting that they feel in control of the information

they reveal (Acquisti & Gross, 2006). However, to the best of our knowledge, this paper is the first to demonstrate that provision of control can have a paradoxical effect: Providing users of modern information-sharing technologies with more granular privacy controls may lead them to share more sensitive information with larger, and possibly riskier, audiences.

BACKGROUND AND HYPOTHESES

In the privacy literature, 'control' is construed as instrumental to privacy protection – so much so that privacy itself is often defined as the control over personal information flows (e.g., Kang, 1998; Solove, 2006; Tavani & Moor, 2001). To understand the paradox of control, however, a distinction must be drawn between the *release* of personal information (the action of willingly sharing some private information with a set of recipients), *access* to it, and *usage* by others. Disclosure – releasing personal information – is a necessary precondition for the access, use, and potential misuse of personal information by others. However, the actual costs associated with the release of personal information depend on whether other people actually access the information, and, if so, what they do with the information they access. Like the proverbial tree that falls in a forest, a violation of privacy on the Internet requires more than the posting of information: someone has to actually access and use it. For example, Facebook provides a strong feeling of control, because users can change every detail of their default privacy settings, including what type of information will be available to whom. However, users have very little control over the way in which information, once posted, will be used by a third-party application or by their friends. The third-party application could, for instance, use that information to send invasive targeted advertising to the user, or perhaps for price discrimination (Acquisti & Varian, 2005); a friend could post the information somewhere else, making it accessible to unintended third parties.

Logically, the aspect of control that should be most relevant for a decision to reveal information is control over the usage of information, since once information is released this is the form of control that would enable the divulger to limit any negative consequences. That is, logically people should ask themselves: "If I release the information, what is likely to happen to it?" However, research on bounded rationality (Simon, 1982) and level-k thinking (Crawford, 2003; Ho, Camerer, & Weigelt, 1998; Nagel, 1995; Stahl & Wilson, 1994) shows that people often fail to engage in conditional thinking. To the degree that people fail to do so (i.e., *not* asking themselves the question of what might happen to information if they were to release it), they may focus on the most proximate level of control they have – control over release – at the expense of contemplating the actual consequences of information access and usage. Based on this logic, we predicted that people who had not yet decided whether to reveal information would fail to appreciate that control over access and usage is much more relevant than control over release once information has, in fact, been released.

Moreover, research on limited attention (e.g., Broadbent, 1957, 1982; Dukas, 2004; Johnston & Dark, 1986; Lachter, Forster, & Ruthruff, 2004; Neisser, 1967; Pashler, 1998) suggests that the human cognitive system has limited capacity and is unable to process the vast amount of information that it constantly receives. Information processing has to be selective, so when emphasis is put on a primary task, attention to secondary tasks tends to decline (Kahneman, 1973). This logic suggests that focusing people's attention on information about their level of control (or lack thereof) over release of personal data is likely to distract them from the lack of control they have over the usage that other people make of the information.

Finally, the release of information, and the choice of recipients originally intended to access it, is what people have control over, whereas actual usage involves actions by others. If

people tend to overestimate the importance of their own actions relative to others', a phenomenon documented by a large body of psychological research on 'egocentrism' and perspective taking (e.g., Gilovich, Medvec & Savitsky, 2000; Epley, Keysar, Van Boven & Gilovich, 2004; Galinsky, 2002; Galinsky, Magee, Inesi & Gruenfeld, 2006), they will, again, focus on their perceived control over release and access, rather than the more compelling source of risk introduced by the uncontrollable actions of others.

In three studies we test whether these three effects can produce the kind of 'control paradox' defined in the introduction. Studies 1 and 2 manipulate control over the *release* of information, and Study 3 manipulates control over *access* (but not usage) of information. The three studies show that perceived control over release or access of personal information can cause people to experience an illusory sense of security and, thus, release more information. Vice versa, lack of perceived control can generate paradoxically high privacy concerns and decrease willingness to disclose, even if the associated risks of disclosure may be lower. In addition, Study 2 tests whether, by focusing scarce attention on issues of control over release, individuals become less sensitive to other normatively relevant factors that serve as cues to objective privacy risks.

All three studies employ a paradigm that is almost ubiquitous in the experimental literature on privacy and information disclosure (e.g., see Joinson, Woodley, & Reips, 2007; Phelps et al. 2000, including most of the 39 studies reviewed in a meta-analysis by Weisband and Kiesler, 1996): they measure concern for privacy by people's propensity to answer personal questions in a survey (see, e.g., Frey, 1986; Singer, Hippler, & Schwarz, 1992).

STUDY 1

Study 1 examines the impact of decreasing control over the release of personal information on willingness to disclose, when this decrease is actually associated with lower probability of access or use of the information by others (and therefore, conditional on disclosure, lower objective benefits, but also lower objective risks). Students at a North American university were recruited to participate in a survey, with the promise of snacks. Participants were invited to become members of a new campus-wide networking website that was supposedly slated to be launched at the end of the semester and populated with profiles automatically created with the information provided during the survey. The survey contained forty questions, which varied in intrusiveness, about the respondent's life in the city and on campus. Intrusiveness was measured in an initial survey of a separate sample of students from the same population. Instructions specified that none of the questions required an answer, but that all answers provided would be part of a profile that would appear on the website, visible to the university community only.

Design

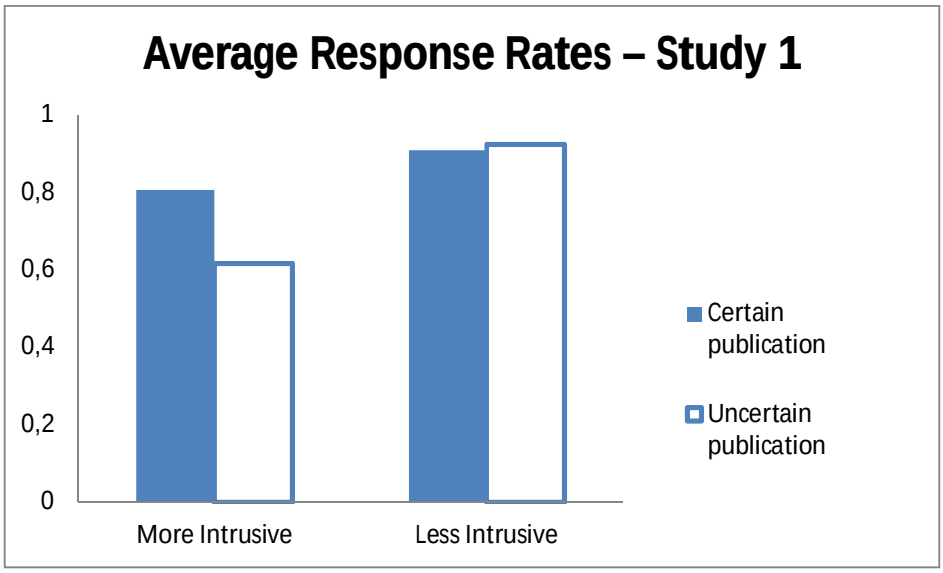
The study was a between-subjects design with two conditions. Participants in the Certain Publication condition were told that a profile would be automatically created for them containing the information they provided, and that this profile would be published online once the website was completed. Participants in the Uncertain Publication condition were told that only half of the profiles created would be randomly selected to be published online. By inserting a random element in the publication process, the Uncertain Publication condition was intended to decrease participants' feeling of control over the public release of their survey answers, while actually

reducing the probability of access by others. According to our hypotheses, the effect of decreased control would reduce willingness to disclose in the Uncertain Publication condition, even though objective costs or risks associated with disclosure were actually lower.

Results

Sixty-seven participants were assigned to the Certain Publication condition, and 65 to the Uncertain Publication condition (overall, 53% female; average age = 21.5, SD = 2.85). Figure 1 shows the average response rate (percentage of questions answered, averaged across participants) by level of intrusiveness of the questions.

Figure 1: Average response rate by type of question in Condition 1 (filled blue, Certain Publication) and in Condition 2 (no-fill blue, Uncertain Publication) – Study 1.



Across conditions, subjects were less likely to answer the more intrusive questions than the less intrusive ones ($t(130) = 11.41$, $p < 0.001$). Supporting our hypotheses, the main effect of control was significant ($F(1,130) = 7.71$, $p < 0.001$). Moreover, as one would expect if control specifically influences concern about privacy, the two-way interaction between condition and

question intrusiveness was significant ($F(1,130) = 32.43$, $p < 0.001$): participants with lower control over information release were significantly less willing to answer personal questions, but especially so for more intrusive questions. The average response rate for intrusive questions was 80.8% in the Certain Publication condition and 61.5% in the Uncertain Publication condition ($t(130) = 4.16$, $p < 0.001$).

A lower response rate in the Uncertain Publication condition could also be attributable to diminished motivation to reveal information when it is less likely that the information will be publicly viewed; halving the probability of publication reduced not just the risks, but also the benefits of disclosure. However, contrary to such an alternative account of the findings, intrusiveness had a negative effect on willingness to reveal. This suggests that participants were motivated to protect their sensitive information. Also, in the presence of diminished motivation, we should have observed lower response rates by subjects in the Uncertain Publication condition to questions that would take more effort to answer. We included in the survey open-ended questions regarding courses attended and enrollment programs to test this interpretation. A regression of aggregate word counts for the open-ended questions failed to reveal any statistically significant difference across the two conditions.

The results of Study 1, therefore, suggest that people respond to manipulations of control over release of personal information in a paradoxical way: Even though lower control implied lower objective risk of accessibility and usage of personal information by others, participants were less willing to disclose if they were provided less control over information release.

STUDY 2

Design

In Study 2, we examined the impact on the propensity to answer privacy-intrusive questions of decreasing participants' perception of control over the release of personal information, while increasing the information's degree of accessibility and potential use by other, potentially more hazardous, recipients. Adopting a 2x2 between-subjects design, Study 2 extended Study 1 by adding a between-subjects manipulation of the accessibility of the information provided. University students, recruited at the same locations as Study 1, answered a shorter version of the same survey. For each of the conditions in Study 1 (Certain versus Uncertain Publication of participants' profiles), new conditions were created that increased accessibility by others: participants read that the website would be accessible either by students only, or by students and faculty members. If one manipulation draws attention to the release of personal information, the other draws attention to its direct accessibility. The survey ended with measures of privacy and accessibility concerns, and a set of manipulation checks regarding perceived control and accessibility of the information provided.

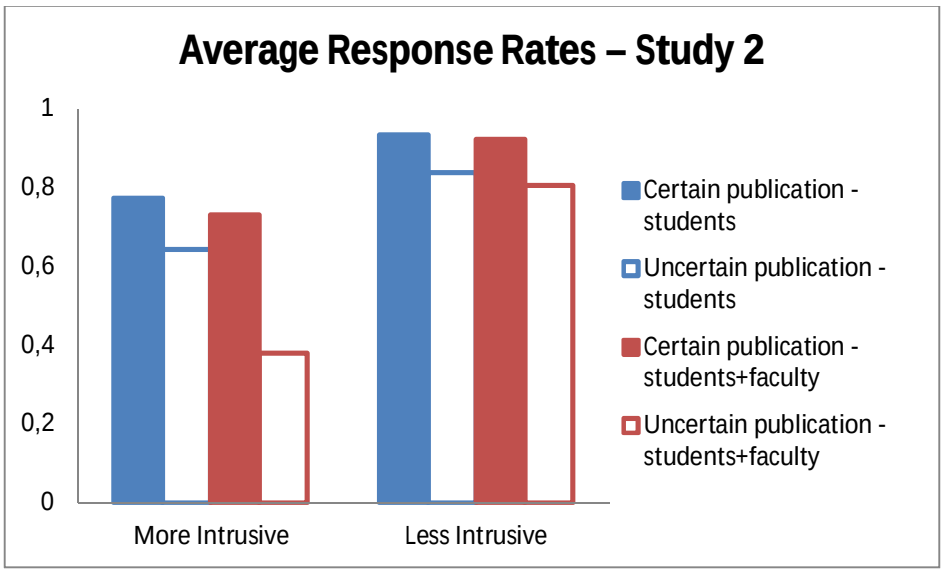
We expected participants' willingness to disclose to be negatively affected by the accessibility of their profiles to faculty members. However, and more to the point, we expected this effect to be dampened in the case of certain publication if, consistently with the limited attention effect discussed in the introduction, participants focused on control over the release of personal information at the expense of its accessibility.

Results

Two-hundred subjects participated in Study 2 (60% female, average age = 21.3, SD = 2.23). Supporting our hypotheses, and replicating the results from Study 1, the main effect of control on question-responding was significant ($F(1,196) = 36.4$, $p < 0.001$). Moreover, similarly to Study 1, there was a significant two-way interaction between control over release and question intrusiveness ($F(1,196) = 15.67$, $p < 0.001$). The main effect of accessibility by faculty was also significant ($F(1,196) = 7.86$, $p < 0.01$), but, as predicted, it was smaller in the case of certain publication, as indicated by the significant interaction of control and accessibility ($F(1,196) = 4.12$, $p < 0.05$). When disclosure was uncertain, participants were less willing to answer intrusive questions if the audience was composed of students and faculty as compared to students only ($t(98) = 3.92$, $p < .001$). This difference was, however, smaller and barely significant when disclosure was certain ($t(98) = .864$, $p = .052$). Reassurances about their control over privacy seemed to decrease participants' attention to issues of accessibility and usage.

Manipulation checks indicate that our manipulation of control over information release was effective, as participants perceived lower control if the publication of their profile was uncertain ($t(198) = -15.53$, $p < 0.001$). The main effects of control over release and accessibility were significant: when asked about accessibility concerns, participants were found to be more concerned if the publication of the profile was uncertain ($F(1,196) = 96.51$, $p < 0.001$), and if the networking website was described as accessible to both students and faculty ($F(1,196) = 15.79$, $p < 0.001$). The interaction was not significant. However, accessibility concerns were actually higher if the publication was uncertain and the audience was composed of students only, than in the case where the publication was certain and the audience included both students and faculty ($t(98) = 3.74$, $p < 0.001$).

Figure 2: Average response rate by type of question in Condition 1 (filled blue; Certain Publication, Students accessibility), Condition 2 (no-fill blue; Uncertain Publication, Students accessibility), Condition 3 (filled red; Certain Publication, Students + Faculty accessibility) and Condition 4 (no-fill red; Uncertain Publication, Students + Faculty accessibility) – Study 2.



Similar results were found for privacy concerns; participants reported higher concerns if the publication of the profile was uncertain ($F(1,196) = 215.36$, $p < 0.001$), and if the networking website was accessible to both students and faculty ($F(1,196) = 5.01$, $p < 0.05$). The interaction was not significant; however, reported privacy concerns were higher if the publication was uncertain and the audience was composed of students only, than in the case where the publication was certain and the audience included both students and faculty ($t(98) = 8.50$, $p < 0.001$). This may suggest that privacy concerns mediate the effect of control on willingness to disclose – a conjecture that we test in Study 3.

Overall, Study 2 supports the central idea that privacy concerns are affected by control over release of personal information, and that reassurances about control over release can distract people from concerns about potentially more hazardous accessibility.[3]

STUDY 3

In contrast to Studies 1 and 2, Study 3 tested the impact of providing participants with more, rather than less, control over the release of their information, and with more control over the actual accessibility of disclosed information. Study 3 also extended the previous studies by testing whether the effect of control applies even to the disclosure of information that could be used to personally identify the divulger, which would significantly heighten the objective risk of privacy violations (Sweeney, 1997). Finally, in Study 3 we tested whether privacy concerns mediated the effect of the experimental manipulations on willingness to disclose.

Using similar recruitment methods as the previous studies, participants were invited to take a survey on 'ethical behaviors.' The survey consisted of ten yes/no questions regarding more or less sensitive behaviors, such as stealing, lying, and consuming drugs. Perceived intrusiveness of the questions was established following the same procedure used in the previous studies.

Participants were informed that none of the questions required an answer, and that the researchers intended to publish the results of the study – including participants' anonymous survey answers – in a Research Bulletin. No detail was given as to whom this Bulletin would be accessible, which was a constant feature across all conditions.

[3] Consistent results were obtained from a similar study, in which accessibility was manipulated telling students that their online profile would be accessible to either members of their own university only, or to members of both their own university and of another, larger university in the same neighborhood. The results are available from the authors.

Design

The study was a non-factorial between-subjects design with four conditions, characterized by increasing control over release of personal information.

In the Implicit Control condition, participants read that by answering a question they would automatically give the researchers permission to publish the answer provided in a Research Bulletin. Participants could decide *not* to answer any question, and therefore deny the researchers the ability to publish their answers, but, unlike the other conditions, there was no explicit mention of the existence of such control.

In the Explicit Aggregate Control condition, before answering the ten questions on ethical behaviors, participants were asked to check a box if they agreed to give the researchers permission to publish *all* their answers among the results of the study. The default option was that the answers would not be published.

In the Explicit Granular Control condition, for each individual question participants were asked to check a box, next to the question, to signal that they were willing to grant publication permission of their answer to that specific question. The default option was that the answers would not be published. This condition emulates several Web 2.0 services, such as blogs and online social networks, which provide users with granular control on what to publish online. [4]

Finally, the Explicit Aggregate Control with Demographics condition was identical to the Explicit Aggregate Control condition, but asked for permission to publish demographic information (in all of the other conditions, participants read that the demographic information

[4] The original study included one additional condition, similar to Condition 3, but with the default consisting of the answers being granted publication. The purpose of that condition was to make sure that default effects were not the main driver for allowing (or not) the publication of the answers. The results are consistent with those presented in the manuscript, and are available from the authors.

they provided would not be published). Participants could click on separate publication permission boxes for gender, age, and country of birth. Releasing this type of personal information is objectively riskier than releasing only answers to ethical behaviors, as it greatly increases the risk that participants could be identified.

We expected to see larger willingness to disclose as the granularity of privacy controls increased, especially for more intrusive questions. Therefore, we predicted that willingness to disclose would be lowest in the Implicit Control condition and highest in the Explicit Granular Control condition, with the remaining conditions in between. Consistent with a control paradox, we predicted that privacy concerns would be soothed by the existence of *explicit control* over access, leading to greater public disclosure of personal information.

Similarly to Study 2, the survey ended with a measure of privacy and accessibility concern and a set of manipulation checks regarding perceived control.

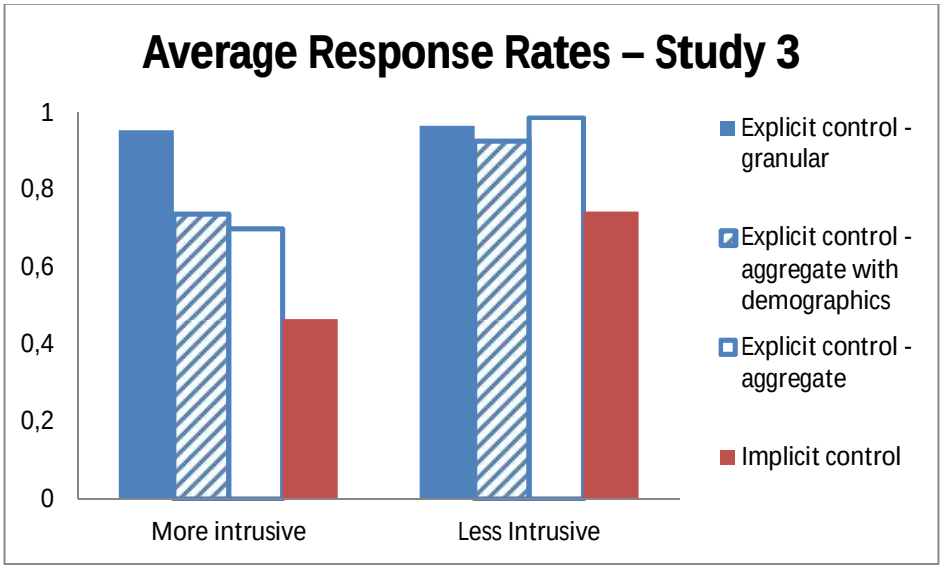
Note that for participants with implicit control, answering a question implied the publication of the corresponding answer, while participants in all other conditions could decide to provide an answer but, when explicitly asked, grant no publication permission. Given this setup, to meaningfully compare results across all conditions, we compared the level of positive responses in the control condition to responses that participants not only provided but also consented to be published.[5]

[5] The results obtained considering response rates (the DV used for the previous studies) are similar to those obtained for publication rates, and are available from the authors.

Results

One-hundred and thirty-four subjects participated in Study 3 (50% female, average age = 21.9, SD = 2.72).

Figure 3: Average response rate by type of question in Condition 1 (filled red; Implicit Control on publication), in Condition 2 (no-fill blue; Explicit Control, Aggregate), in Condition 3 (filled blue; Explicit Control, Granular), and in Condition 4 (striped blue; Explicit Control, Aggregate with Demographics) – Study 3.



All participants in the Explicit Aggregate Control conditions, with and without demographics, checked the publication permission box, thus allowing the public release of their answers. Moreover, all participants in the Demographics condition granted permission to publish all three demographic items – which dramatically increases their identifiability (Sweeney, 1997). This striking result suggests that, as long as people perceive control over the decision to publish personal information and the audience to whom access will be granted, they will indeed decide to publish it, even if the objective risks associated with disclosure increase dramatically. The main effect of control over information release was significant ($F(3,130) = 33.53$, $p < 0.001$): Figure 3 shows that willingness to disclose increases as the level of control increases from implicit to

explicit-aggregate and to explicit-granular. In addition, consistent with the idea that control influences concern about privacy, the two-way interaction between condition and question intrusiveness was significant ($F(3,130) = 11.98$, $p < 0.001$). Supporting our hypothesis that perceived control decreases people's sensitivity to privacy violations, voluntarily revealing demographic information in the Demographics condition did not affect willingness to answer sensitive questions, even though the objective risk of disclosure was higher.

These results suggest that reported privacy concerns should mediate the effect of actual control (dummy variables representing all conditions with explicit control) on willingness to disclose. To test this, we included our measure of privacy concern in a mediation analysis (Table 1). We conducted an OLS regression using a bootstrapping technique (Preacher & Hayes, 2008).

The total effect of actual control on willingness to disclose was positive and significant, as the coefficients on all three dummies were significantly larger than zero (Model 1: $\beta_2 = .25$, SE = .04, $t(130) = 6.47$, $p < .001$; $\beta_3 = .36$, SE = .04, $t(130) = 9.84$, $p < .001$; $\beta_4 = .24$, SE = .04, $t(130) = 6.25$, $p < .001$). Privacy concern correlated negatively with actual control (Model 2: $\beta_2 = -.82$, SE = .43, $t(130) = -1.91$, $p = .06$; $\beta_3 = -2.11$, SE = .42, $t(130) = -5.06$, $p < .001$; $\beta_4 = -1.82$, SE = .43, $t(130) = -4.26$, $p < .001$). Accounting for privacy concerns, the relationship between actual control and willingness to disclose weakened (Model 3: $\beta_2 = .22$, SE = .04, $t(129) = 6.08$, $p < .001$; $\beta_3 = .30$, SE = .04, $t(129) = 7.87$, $p < .001$; $\beta_4 = .18$, SE = .04, $t(129) = 4.78$, $p < .001$). A bootstrap analysis revealed that the 95% bias-corrected confidence interval for the size of the indirect effects excluded zero, which suggested a significant indirect effect (MacKinnon, Fairchild, & Fritz, 2007; Preacher & Hayes, 2004).

Table 1. Mediation analysis – Study 3. Model 1 and 3: Dependent Variable is Average Response Rate. Model 2: Dependent Variable is Privacy Concerns. Standard errors in brackets. 95% bootstrapped CI in squared brackets. *indicates significance at 10% level **indicates significance at 5% level. ***indicates significance at 1% level.

	Model 1	***Model 2***	***Model 3***
Average Response Rate	-	-	
Condition 2	.247*** (.038)	-.822* (.43)	.223*** (.034) [.0002, .0460]
Condition 3	.364*** (.037)	-2.114*** (.418)	.302*** (.038) [.0245, .1074]
Condition 4	.236*** (.038)	-1.818*** (.427)	.183** (.038) [.0224, .0922]
Privacy Concerns	-	-	-.029*** (.007)
	N = 134 F(3,130) = 33.53	N = 134 F(3,130) = 10.51	N = 134 F(4,129) = 32.06

This study shows that, paradoxically, participants were more likely to allow the publication of information about them, and more likely to disclose *more* information of a sensitive nature, as long as they were *explicitly,* instead of *implicitly,* given control over its publication. Participants in the Implicit Control condition could avoid publication by not answering questions; but participants in the other conditions, who had an explicit option to publish their answers and determine the level of their accessibility, felt less privacy concerned and thus became more likely to not just *answer*, but also allow the *publication* of their answers. It was not the publication of personal information *per se* that modulated privacy concerns, but rather the explicit perceived control over it.

DISCUSSION

Three experiments provide empirical evidence that perceived control over release plays a critical role in (over)sharing personal information, relative to the objective risks associated with information access and usage by others. In Study 1, participants responded to manipulations that decreased control over information release, even though risks associated with information access and use by others were in fact decreased. In Study 2, control over release distracted participants from concerns about potentially more hazardous accessibility. In Study 3, participants given explicit control over the release and accessibility of their personal information revealed more, even exposing themselves to higher risks of identifiability.

Our findings introduce a novel scenario in the scholarly literature on privacy and control, where it has been conventional wisdom that control over personal information either implies (Culnan, 1993; Elgesem, 1996; Fried, 1984; Lessig, 2002; Miller, 1971; Smith, Milberg, & Burke, 1996; Westin, 1967), or at most does not negatively affect (Laufer & Wolfe, 1977; Tavani & Moor, 2001) privacy protection. Our results show that 'more' control can sometimes lead to 'less' privacy in the sense of higher objective risks associated with the disclosure of personal information. In other words: our results provide evidence that control over personal information may be a necessary (in ethical or normative terms) but not sufficient condition for privacy protection.

Notice that our argument does not posit that people *should* be concerned about their privacy, or that they *have to* disclose less in order to achieve higher utility or satisfaction. While recent research on regrets associated with online information sharing does indicate that, at times, people feel they revealed too much (Wang et al., 2011), in our studies, and indeed most

situations, there is no objective standard for determining whether participants revealed too little or too much. To document that privacy-related behavior is suboptimal, therefore, we show that people change their propensity to disclose in response to non-normative factors (such as whether they have explicit or implicit control over publication) and fail to change their disclosure behavior (or even change in the wrong direction) in response to normative factors (such as whether they can be personally identified).

The conventional wisdom that control is an essential component of privacy is so ubiquitous that 'control' has become a code-word employed both by legislators and government bodies in proposals for enhanced privacy protection, and by data holders and service providers to deflect criticisms regarding the privacy risks borne by data subjects. For instance, Facebook's CEO Mark Zuckerberg has repeatedly stressed the role of privacy controls as instruments to have "more confidence as you share things on Facebook,"[6] while both Senator Kerry's bill proposal and the recent Federal Trade Commission's Privacy Report focus on giving users more (privacy) control.[7] In fact, numerous government and corporate entities in the United States have advocated self-regulatory 'choice and consent' models of privacy protection that, essentially, rely on users' awareness and control.

The argument is appealing; users *do* want more control over how their information is collected and used (Consumer Reports National Research Center, September 2008, http://www.consumersunion.org/pub/core_telecom_and_utilities/006189.html). However, higher levels of control may not always serve the ultimate goal of enhancing privacy protection. The paradoxical policy implication of these findings is that the feeling of security conveyed by the

[6] "Giving you more control," posted by Mark Zuckerberg on October 10, 2010, available at http://www.facebook.com/blog.php?post=434691727130.
[7] See http://kerry.senate.gov/press/release/?id=223b8aac-0364-4824-abad-274600dffe1c and http://www.ftc.gov/os/2010/12/101201privacyreport.pdf.

provision of fine-grained privacy controls may lower concerns regarding the actual accessibility and usability of information, driving those provided with such protections to reveal more sensitive information to a larger audience.

APPENDIX

Survey questions by level of intrusiveness – Study 1

Rating	Questions
Very intrusive	Q7: Email address Q8: Home address Q9: Phone number Q28: Have you ever had a sexual relationship with somebody other than your partner without their knowledge or consent? Q34: Have you ever cheated for homework/projects (e.g. copy, plagiarize) or on an exam? Q35: Have you ever seen someone else cheating? Q36: If so, did you inform the instructor?
Moderately intrusive	Q4: Date of birth Q5: Age (in years) Q19: If so, which group or groups are you a member of? Q20: How many of the people you know in [city name] do you consider close friends? Q27: Do you have a girlfriend/boyfriend? Q29: Where do you live? (University housing, Private housing-alone, Private housing-shared) Q30: Have you ever had troubles with your roommates? Q31: Would you like to move somewhere else?
Not at all intrusive	Q1: First name, Middle name Q2: Last name Q3: Gender Q6: Country of birth Q10: Do you have a Facebook profile? Q11: How long have you been in [city name]? Q12: On a scale from 1 (not at all) to 10 (very much), how do you like the city overall? Q13: How happy are you here? Q14: Do you do any sport? Q15: If so, which sport do you do? Q16: Do you do any sport on campus? Q17: How would rate the sport facilities offered on campus? Q18: Are you a member of any group/community/fraternity/sorority? Q21: How many of those are students at [university name]? Q22: How many are students at other universities in [city name]? Q23: Do you enjoy spending your spare time with your friends much more/ with your friends more/ alone more/ alone much more/ with your friends just as much as alone? Q24: Is your family in [city name]? Q25: How often do you see your family? Q26: Are you single or married? Q32: What program are you in? (e.g.: Undergrad Psychology, Grad Math) Q33: Which courses are you taking at the moment? Q37: How would you rate the quality of the education you are receiving on a scale from 1 (very bad) to 5 (very good)? Q38: Do you think it will make you competitive on the job market? Q39: How many hours a day do you spend studying? Q40: Are you working at the same time?

Survey questions by level of intrusiveness – Study 2 (subset of questions in Study 1)

Rating	Questions
Very intrusive	Q7: Email address Q8: Home address Q9: Phone number Q28: Have you ever had a sexual relationship with somebody other than your partner without their knowledge or consent? Q34: Have you ever cheated for homework/projects (e.g. copy, plagiarize) or on an exam? Q35: Have you ever seen someone else cheating? Q36: If so, did you inform the instructor?
Moderately intrusive	Q4: Date of birth Q5: Age (in years) Q19: If so, which group or groups are you a member of? Q20: How many of the people you know in [city name] do you consider close friends? Q27: Do you have a girlfriend/boyfriend? Q29: Where do you live? (University housing, Private housing-alone, Private housing-shared) Q30: Have you ever had troubles with your roommates? Q31: Would you like to move somewhere else?
Not at all intrusive	Q1: First name, Middle name Q2: Last name Q3: Gender Q6: Country of birth Q10: Do you have a Facebook profile? Q18: Are you a member of any group/community/fraternity/sorority? Q21: How many of those are students at [university name]? Q22: How many are students at other universities in [city name]? Q24: Is your family in [city name]? Q25: How often do you see your family? Q26: Are you single or married? Q32: What program are you in? (e.g.: Undergrad Psychology, Grad Math) Q33: Which courses are you taking at the moment? Q39: How many hours a day do you spend studying? Q40: Are you working at the same time?

Exit questions – Study 2

M1: Have you understood how your answers will be used? Please describe.

Accessibility concerns – M2: As you answered the questions in this survey, how concerned were you about who would access the information you provided? (Not at all – Very much) Please briefly explain.

Privacy concerns – M3: How concerned were you about your privacy as you answered the questions in this survey? (Not at all – Very much) Please briefly explain why you felt that way.

Perceived control – M4: Do you think you were given enough control on whether your answers would be published on your profile? (By control we refer to whether you felt you could decide what information would be published or not) (No control at all – Complete control) Please briefly explain.

Perceived accessibility – M5: Who will be able to access the networking website, and therefore view your profile? ([OWN UNIVERSITY] members / [OWN UNIVERSITY AND OTHER UNIVERSITY] members / Everybody on the Internet / I don't know)

Perceived accessibility – M6: Do you think there is a possibility that someone else could view your profile? (Yes / No / I don't know) Please briefly explain.

Survey questions – Study 3

Demographics
D1: Age
D2: Gender
D3: Country of birth
D4: Email address

Questions on ethical behaviors – Study 3

Rating	Questions
Very intrusive	Q2: Have you ever been fired by your employer? Q3: Have you ever stolen anything (e.g.: from a shop, a person)? Q4: Have you ever used drugs of any kind (e.g.: weed, heroin, crack)? Q6: Have you ever had cosmetic surgery? Q8: Have you ever had sex in a public venue (e.g.: restroom of a club, airplane)?
Moderately intrusive	Q10: Do you have any permanent tattoos?
Not at all intrusive	Q1: Are you married? Q5: Have you ever lied about your age? Q7: Have you ever done any kind of voluntary service? Q9: Have you ever made a donation to a non-profit organization?

Exit questions – Study 3 (MC4 differed across conditions)

MC1: Have you understood how your answers will be used? Please describe.

Accessibility concerns – MC2: In answering the questions in the previous page, were you concerned about the publication of the information provided? Please briefly explain.

Perceived control – MC3: Do you think you were given enough control on whether your answers would be published among the results of the study? (By control we refer to whether you felt you could decide what would be published or not).

Perceived control – MC4

Condition 1: How did you feel about the fact that, for all the questions you actually answered, you could not control their publication? If you didn't feel neither one way or the other, please click on the middle choice. (Annoyed – Pleased, Powerless – Empowered, Frustrated – Calm, Controlled – Autonomous, Embarrassed – At ease)

Condition 2: How did you feel about the fact that, for each question you actually answered, you could not control its individual publication? If you didn't feel neither one way or the other, please click on the middle choice. (Annoyed – Pleased, Powerless – Empowered, Frustrated – Calm, Controlled – Autonomous, Embarrassed – At ease)

Condition 3: How did you feel about the fact that, for all the questions you actually answered, you had to check a box to allow their publication? If you didn't feel neither one way or the other, please click on the middle choice. (Annoyed – Pleased, Powerless – Empowered, Frustrated – Calm, Controlled – Autonomous, Embarrassed – At ease)

Condition 4: How did you feel about the fact that, for each non demographic question you actually answered, you could not control its individual publication? If you didn't feel neither one way or the other, please click on the middle choice. (Annoyed – Pleased, Powerless – Empowered, Frustrated – Calm, Controlled – Autonomous, Embarrassed – At ease)

Privacy concerns – MC5: How concerned were you about your privacy as you answered the questions in this survey?

If you didn't feel neither one way or the other, please click on the middle choice. (Not at all – Very Much) Please briefly explain why you felt that way.

MC6: Do you think that your email address and/or your demographic information will be published among the results of the study? Please briefly explain.

Survey Instructions

Study 1

Instructions in the Certain Publication condition

Please read these instructions carefully before you move on. This is not the usual yada-yada. The information you provide will appear on a profile that will be automatically created for you. The profile will be published on a new [university name] networking website, which will only be accessible by members of the [university name] community, starting at the end of this semester. The data will not be used in any other way. NO QUESTION/FIELD REQUIRES AN ANSWER. Did you understand these instructions? If so, click on Next.

Instructions in the Uncertain Publication condition

Please read these instructions carefully before you move on. This is not the usual yada-yada. The information you provide will appear on a profile that will be automatically created for you. Half of the profiles created for the participants will be randomly picked to be published on a new [university name] networking website, which will only be accessible by members of the [university name] community, starting at the end of this semester. The data will not be used in any other way. NO QUESTION/FIELD REQUIRES AN ANSWER. Did you understand these instructions? If so, click on Next.

Study 2

Instructions in the Certain Publication – Students condition

Please read these instructions carefully before you move on. This is not the usual yada-yada. The information you provide will appear on a profile that will be automatically created for you. The profile will be published on a new [university name] networking website, which will only be accessible by students of [university name], starting at the end of this semester. The data will not be used in any other way. NO QUESTION/FIELD REQUIRES AN ANSWER. Did you understand these instructions? If so, click on Next.

Instructions in the Uncertain Publication – Students condition

Please read these instructions carefully before you move on. This is not the usual yada-yada. The information you provide will appear on a profile that will be automatically created for you. Half of the profiles created for the participants will be randomly picked to be published on a new

[university name] networking website, which will only be accessible by students of [university name], starting at the end of this semester. Limited server capacity in this launching phase of the website is the reason why only half of the profiles will be published. The data will not be used in any other way. NO QUESTION/FIELD REQUIRES AN ANSWER. Did you understand these instructions? If so, click on Next.

Instructions in the Certain Publication – Students + Faculty condition

Please read these instructions carefully before you move on. This is not the usual yada-yada. The information you provide will appear on a profile that will be automatically created for you. The profile will be published on a new [university name] networking website, which will only be accessible by students and faculty of [university name], starting at the end of this semester. The data will not be used in any other way. NO QUESTION/FIELD REQUIRES AN ANSWER. Did you understand these instructions? If so, click on Next.

Instructions in the Uncertain Publication – Students + Faculty condition

Please read these instructions carefully before you move on. This is not the usual yada-yada. The information you provide will appear on a profile that will be automatically created for you. Half of the profiles created for the participants will be randomly picked to be published on a new [university name] networking website, which will only be accessible by students and faculty of [university name], starting at the end of this semester. Limited server capacity in this launching phase of the website is the reason why only half of the profiles will be published. The data will not be used in any other way. NO QUESTION/FIELD REQUIRES AN ANSWER. Did you understand these instructions? If so, click on Next.

Study 3

Instructions in the Implicit Control condition

Please read the following instructions carefully, as they are important.

In the following pages, you will be asked a number of questions related to ethical behavior.

The researchers would like to publish the answers you will provide in the following pages on their Research Bulletin, among the results of the study.

Please notice that the answers to the demographic questions that you provided in the previous page will NOT be published.

All answers are voluntary. By answering a question, you agree to give the researchers permission to publish your answer.

Instructions in the Explicit Aggregate Control condition

Please read the following instructions carefully, as they are important.

In the following pages, you will be asked a number of questions related to ethical behavior.

The researchers would like to publish the answers you will provide in the following pages on their Research Bulletin, among the results of the study.

Please notice that the answers to the demographic questions that you provided in the previous page will NOT be published.

All answers are voluntary. In order to give the researchers permission to publish your answers to the questions, you will be asked to check a box in the following page.

Instructions in the Explicit Granular Control condition

Please read the following instructions carefully, as they are important.

In the following pages, you will be asked a number of questions related to ethical behavior.

The researchers would like to publish the answers you will provide in the following pages on their Research Bulletin, among the results of the study.

Please notice that the answers to the demographic questions that you provided in the previous page will NOT be published.

All answers are voluntary. In order to give the researchers permission to publish your answer to a question, you will be asked to check the corresponding box in the following page.

Instructions in the Explicit Aggregate Control condition with Demographics

Please read the following instructions carefully, as they are important.

In the following pages, you will be asked a number of questions related to ethical behavior.

The researchers would like to publish the answers you will provide in the following pages on their Research Bulletin, among the results of the study.

All answers are voluntary. In order to give the researchers permission to publish your answers to the questions, you will be asked to check a box in the following page.

Please notice that the answers to the demographic questions that you provided in the previous page will NOT be published without your explicit agreement: you will be asked permission to publish those answers separately.

Chapter 3

Differential depreciation of information with positive and negative valence: The role of diagnosticity

with Joachim Vosgerau and Alessandro Acquisti

Abstract

Negative events in the future are discounted less than positive events, and people adapt slower to negative events in the past than positive events. We investigate whether this 'bad has a longer lasting impact than good' principle also holds for impression formation. We hypothesize that diagnostic past behaviors of a person have a longer lasting impact on others' impression of, and behavior towards that person, than her non-diagnostic past behaviors. In experiment 1, we show that a negative diagnostic (immoral) behavior has a longer lasting impact on impression formation than a positive non-diagnostic (moral) behavior. In experiment 2, we demonstrate that past diagnostic behaviors have a longer lasting impact on others behavior towards the target person, but only when the person's past behaviors were voluntary. Finally, in experiment 3 we show that when past positive behaviors are diagnostic, they have a longer lasting impact than past negative behaviors.

INTRODUCTION

"Glory is fleeting, but obscurity is forever."
- Napoleon Bonaparte (1769-1821)

In January 2007, two Ottawa employees of a grocery store chain were fired after their employer found a Facebook group where, two months earlier, the employees made admissions of theft on the groups' message board (CNEWS, Canada, Jan. 17th 2007).

Many graduating students are aware of the pitfalls of having documented their lives too extensively on social network sites, and suspend or delete their profiles prior to going on the job market. They do so because they correctly anticipate that displaying their past negative behaviors – such as admitting theft – can have much more deleterious consequences than displaying their past positive behaviors can bear positive consequences – as, for example, caring for a friend struggling with cancer. In virtually all domains of human behavior including judgment, information processing, and learning, bad information has been shown to have a stronger impact than good information (for two comprehensive reviews see Rozin & Royzman, 2001, and Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). In order to avoid the consequences of the stronger impact of negative behaviors, however, is it necessary to delete one's entire profile, or would it suffice to just delete the most recent entries without investing time and effort into clearing one's whole online history and traces? In other words, is a 'negative' behavior displayed 5 years ago still more damaging than a few months old 'positive' behavior is self-enhancing? The current research investigates this question.

Two distinct literature streams have documented differential changes of the impact of negative/positive information over time: research on time preferences and on hedonic adaptation. The behavioral economics literature on time preferences has shown that the more an event lies in

the future the less it impacts current valuations and choices (discounting of the future). More importantly, not all events are discounted equally. Many studies have demonstrated that small amounts of money are discounted faster than large amounts of money (the so-called 'magnitude-effect'), and gains are discounted faster than equivalent losses (the 'sign-effect;' for an overview of these studies see Frederick, Loewenstein, & O'Donoghue, 2002).

Paralleling these findings, several studies in the literature on hedonic adaptation suggest that people adapt slower to negative events than to positive events. For example, Brickman, Coates, and Janoff-Bulman (1978) interviewed people who had won a lottery a year ago, people who had been paralyzed in an accident a year ago, and a control group. While the lottery winners were no happier than the control group, the accident victims were significantly less happy. A similar conclusion has been drawn from research on child abuse and sexual abuse. Even if the abuse occurred only once or twice, it produces long-lasting harmful consequences such as depression, relationship problems, revictimization, and sexual dysfunction, (Cahill, Llewelyn, & Pearson, 1991; Fleming, Mullen, Sibthorpe, & Bammer, 1999; Silver, Boon, & Stones, 1983; Styron & Janoff-Bulman, 1997; Weiss, Longhurst, & Mazure, 1999). In contrast, no positive experiences have been documented to date that have equally long-lasting effects. Finally, correlational diary studies suggest that daily bad events have a longer lasting impact on mood, self-esteem, anxiety, and well-being than daily positive events (for a detailed overview see Baumeister et al., 2001).

While both streams of literature – time preferences and hedonic adaptation – provide empirical evidence that negative information has a longer lasting impact than positive information, the studies in the two research streams differ in their time orientation. Studies on time discounting look at how future gains and losses influence current valuations and choices,

whereas studies on habituation investigate how past negative/positive events affect current wellbeing of the experiencer.

In the current paper, we investigate how a person's past negative behaviors compared to her past positive behaviors affect others' impressions of, and behavior towards that person, depending on when these behaviors occurred in time. Similar to studies of time-discounting, this research looks at whether negative or positive information is depreciated more over time, but like research on hedonic adaptation, it focuses on information from the past rather than the future.

When being presented with information about a target person, people automatically form an impression of that person. These impressions not only include a general evaluation of the target person (Zajonc, 1980), but also evoke trait concepts and stereotypes that activate corresponding behaviors (Bargh, Chen, Burrows, 1996). Baumeister et al. (2001) have argued that it is adaptive to give negative behaviors more weight than positive behavior in the impression formation process, because negative past behaviors may signal negative traits of the target person that – if ignored – could have more severe consequences than ignoring past positive behaviors signaling positive traits. This evolutionary account of why the impact of negative information should decline less over time than the impact of positive information, however, does not specify the psychological process that governs the weight given to each type of information over time.

We hypothesize that information about a person will depreciate less over time the more diagnostic it is for predicting the person's traits. Following Reeder and Brewer's (1979) model of dispositional attribution, immoral behaviors are more diagnostic than moral behaviors (Skowronski & Carlston, 1987; 1989). For example, stealing once is enough for a person to be

labeled a thief, whereas a person needs to behave always morally to be called a person of high integrity. For judgments of ability, in contrast, positive behaviors are more diagnostic than negative behaviors. Winning a chess tournament once is enough to judge a person as intelligent, whereas a person needs to behave stupidly most of the time to be judged as unintelligent. If diagnosticity of past behaviors drives the extent to which these behaviors depreciate over time, we would expect *less* depreciation of negative than positive behaviors in the *domain of morality*, but *more* depreciation of negative than positive behaviors in the *domain of ability*.

Another hypothesis can be derived from our diagnosticity explanation. Activities that a person engaged in voluntarily should be diagnostic for predicting that person's intentions, but involuntary actions should be non-diagnostic of her intentions. Consequently, differential depreciation of past behaviors should only be observed if the past behaviors were performed voluntarily.

In this paper, we test our diagnosticity explanation in three experiments. Following the guidelines outlined in Simmons, Nelson, and Simonsohn (2011), we report all measures that were used in each experiment. All samples were convenience samples, sample sizes were set to a minimum of N = 30 per experimental cell before data collection.

EXPERIMENT 1

Experiment 1 was designed to test whether negative behaviors would depreciate less than positive behaviors in the morality domain. The experiment employed a 2 (valence: positive vs. negative behavior) x 2 (time: recent versus far past) between-subjects design, with an additional control condition.

Participants and Procedure

One hundred sixty-eight students (80 female, M_{age} = 21.7, *SD* = 2.49) at Carnegie Mellon University were recruited for a study on impression formation. Students were approached by the experimenter in a popular indoor gathering place on campus, and asked to complete a 5-minute online survey in exchange for candies. Students who agreed to participate were given a laptop and randomly assigned to one of the five experimental conditions.

Participants read background information about a target person, such as place of birth, schooling, where s/he got her/his driving license, where s/he went to college, her/his hometown and where s/he moved to, and her/his current job. The background information was ordered by time of occurrence. In the negative valence condition, the background information included the additional information that the person had found a wallet with $10,000 in cash but did not report it to the police to be returned to the owner. In the positive valence condition, the person was reported as having returned the wallet with the cash to the police. The negative and positive behaviors were described to have occurred either recently (12 months ago) or in the far past (5 years ago). Importantly, time of occurrence of the critical behaviors was manipulated, while order of presentation was held constant. The control condition included only the neutral behaviors. Participants were then asked to answer a couple of comprehension questions, which all of them answered correctly. After the comprehension questions, participants were asked how much they liked the target person and whether they would like to work with this person on seven-point scales with endpoints, “dislike very much”/“don’t want to work with this person” (1) and “like very much”/“want to work with this person” (7), and whether participants would trust the target person on a three-point scale with endpoints, “low trust (1) and “high trust” (3). The study ended with demographic questions.

Results

Dependent variables. The three dependent variables – liking, wanting to work with the target person, and trust – loaded onto a single factor. We multiplied the trust scores by 7/3 to make them comparable to the other two variables, and averaged across all three variables to obtain an overall index of positive/negative impression of the target person (Cronbach's alpha = .92).

Manipulation check. Recall that time (recent vs. far past) was only manipulated in the positive and negative behavior conditions, but not in the neutral behavior condition. To test whether the behavior valence manipulation was successful, we conducted a (positive vs. neutral vs. negative behavior) one-way ANOVA on the overall impression index, which showed that the valence manipulation was effective ($F(2, 165) = 278.85$, $p < .001$). For the average impression ratings and t-tests between the three conditions see table 1.

Table 1: Average impression of the target person as a function of her previous behaviors in experiment 1. Standard deviations are in parentheses.

previous behaviors by target person	impression of target person	*t*-test
1 positive behavior, all others neutral	5.27 (0.71)	$t(98) = 4.73$ $p < .001$
only neutral behaviors	4.63 (0.40)	$t(98) = 14.22$ $p < .001$
1 negative behavior, all others neutral	2.50 (0.80)	

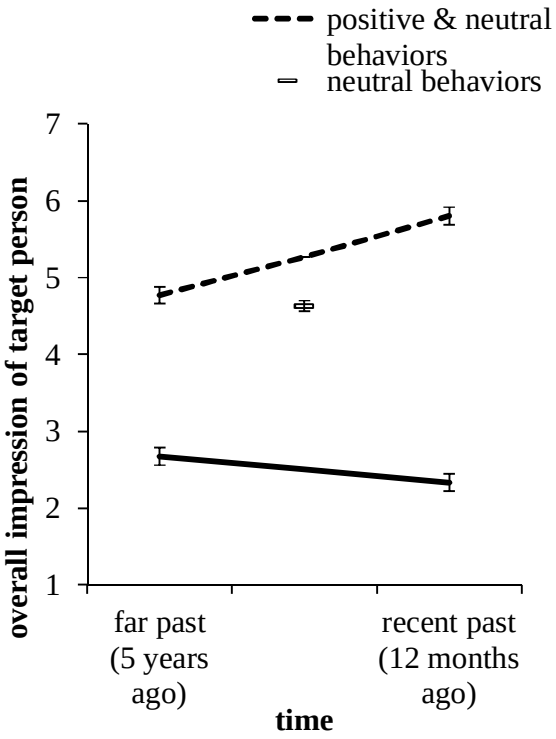
Discounting of negative and positive behaviors. To test whether the impact of the negative behavior would be discounted less over time than the impact of the positive behavior, we mean-centered the impression scores and multiplied them by -1 in the negative behavior conditions. This way, positive and negative behaviors affect the new mean-centered scores in the same direction, and we can compare the slopes over time for positive and negative behaviors. A 2 (positive vs. negative behavior) x 2 (recent vs. far past) ANOVA obtained a main effect for time ($F(1, 132) = 37.72$, $p < .001$), indicating that recent behaviors (12 months ago) had a stronger impact on impressions than behaviors in the far past (5 years old). More importantly, the hypothesized interaction of valence and time was significant ($F(1, 132) = 9.66$, $p = .002$). As predicted, it made little difference whether the negative behavior had occurred 5 years ago ($M = 1.22$, $SD = 0.79$) or 12 months ago ($M = 1.56$, $SD = 0.79$, contrast $t(132) = 2.15$, $p = .034$), but it made a bigger difference whether the positive behavior had occurred 5 years ago ($M = 0.89$, $SD = 0.48$) or 12 months ago ($M = 1.92$, $SD = 0.48$, contrast $t(132) = 6.54$, $p < .001$). Figure 1 displays the raw un-centered impression ratings.

Discussion

The results of Experiment 1 show that people discount past immoral behaviors to a lesser extent than past moral behaviors when forming an overall impression of a person. Notice that this finding cannot be attributed to order effects, as the order of information was kept constant across conditions. While this pattern of results is consistent with diagnostic information (immoral behaviors) being less discounted than less diagnostic information (moral behaviors), alternative explanations exist. First, the immoral behavior was more extreme than the moral behavior (i.e., it led to a more negative impression than the moral behavior led to a positive impression). So, rather than negativity, it may be extremity that caused the lesser discounting of

the immoral behavior. Second, negative social information is more attention-grabbing than positive social information (Pratto & John, 1991), which may have led to less discounting of the negative information. Finally, negative events elicit more causal attribution than positive events (Peeters & Czapinski, 1990), which suggests that people are more likely to infer intentions from immoral than moral behaviors. Note that this explanation is closely related to our explanation of immoral behaviors being more diagnostic of intentions. In experiment 3 we manipulate diagnosticity orthogonally to the valence of the behaviors to test the three alternative explanations against our diagnosticity account.

Figure 1: Overall impression ratings of the target person as a function of behaviors displayed in the recent (12 months) or far past (5 years ago) in experiment 1. Error bars represent +/- 1 standard errors.



EXPERIMENT 2

Experiment 2 was designed to assess behaviors in addition to impressions, and to test whether differential discounting of negative and positive behaviors would occur when the behaviors were performed voluntarily but not when the person involuntarily engaged in the activities. Participants played a dictator game with a – unbeknownst to participants – fictitious player, and had to decide how much real money to allocate to this player. Prior to making their allocation decision, they were shown the players' history of allocations in previous games.

The experiment employed a 2 (valence: generous vs. greedy split; btw-sbj.) x 2 (time: recent vs. far past; btw-sbj.) x 2 (voluntary vs. involuntary; btw-sbj.) x 2(allocation vs. fairness rating; within-sbj.) design, with an additional control condition. We set N =40 per cell and aimed at collecting 360 responses.

Participants and Procedure

Three hundred seventy-three students (200 males, M_{age} = 22.8, SD = 1.32) at Carnegie Mellon University were recruited to play an "allocation game". Students were approached by the experimenter in a popular indoor gathering place on campus, and asked whether they would like to play a game with another player in a virtual game room. Participants logged onto the virtual game room (which we developed and hosted on a university server) on laptops provided by the experimenter.

Prior to starting the game, each participant was told that $1 was provisionally allocated to her and all other (fictitious) players in the virtual game room. Participants would be randomly paired with another player and assigned the role of either "allocator" or "recipient." As allocator, they could decide whether and how to split the money; as recipient, they could only accept the

decision of the allocator. In fact, all participants were assigned the role of allocator. Participants were further told that, on average, allocators keep 70% of the money for themselves and give 30% to the other player (Forsythe, Horowitz, Savin, & Sefton, 1994). This way, we primed participants to consider a 70-30 split a fair allocation.

Before starting the game, participants were asked two comprehension questions, which all 373 participants answered correctly. Participants were then matched with the other (fictitious) player, and informed that said player had already played seven rounds of the game as an allocator. Before deciding how to allocate the money, participants were shown a table summarizing the allocations previously decided by the other player in her previous game. We systematically varied the content of the table to manipulate valence, time, and intentionality of the other player's previous allocations. In the neutral condition, all allocations were close to a 70-30 split. In the generous/greedy conditions, the other player had chosen a generous 50-50/greedy 100-0 split in one of the seven rounds, either in round 2 (far past) or round 6 (recent past). Finally, in the involuntary conditions, participants read that the generous/greedy allocation had been caused by a bug in the program, and the player had no chance to change that allocation.

To provide a realistic feeling of a dynamic game taking place in a crowded virtual game room, rows of the table of previous allocations by the other player appeared sequentially on the screen, simulating the fictitious player making actual allocation decisions with other players before playing with the participant. This design also allowed us to make the temporal order of allocations more salient.

Participants were then asked how they wanted to split the $1 between themselves and the other player (the table with the previous allocations of the other player was still displayed). After

making their allocation decision, participants were asked how fair they perceived the other player to be on a seven point scale with endpoints, “very unfair” (1) and “very fair” (7), and why they chose that allocation. Participants were also asked whether any technical error occurred during the game, and finally their demographics. The final thank-you screen summarized the payoffs, and participants were paid accordingly and debriefed by the experimenter.

Results

Manipulation check. We conducted a (generous vs. fair vs. greedy allocation) MANOVA on the money amount that participants allocated to, and their fairness rating of the other player. The manipulation was successful ($F(4, 740) = 44.87$, $p < .001$). For the average allocations to, and fairness ratings of the other player as a function of her previous allocations see table 2.

Table 2: Average allocations to and fairness ratings of the other player as a function of previous allocations by the other player in experiment 2. Standard deviations are in parentheses.

previous splits by other player	allocation to other player	*t*-tests allocation	fairness rating of other player	*t*-tests fairness
1 generous split (50-50)	\$0.35 (0.12)	$t(207) = 2.06$, $p = .041$	4.43 (0.94)	$t(207) = 3.24$, $p = .001$
all splits fair (~70-30)	\$0.31 (0.10)	$t(204) = 4.44$, $p < .001$	3.93 (0.64)	$t(204) = 6.07$, $p < .001$
1 greedy split (100-0)	\$0.21 (0.14)		2.84 (1.11)	

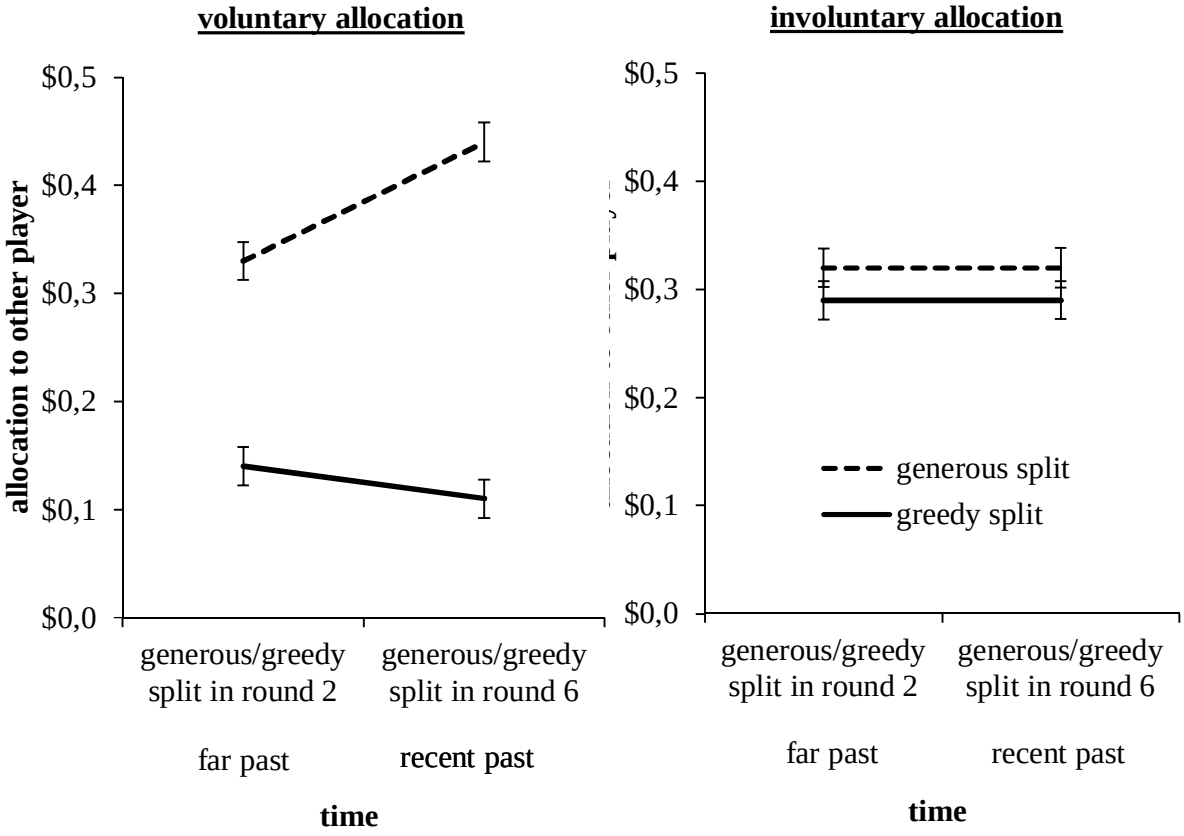
Discounting of previous generous versus greedy allocations. As in the previous experiment, we first mean-centered allocations to, and fairness ratings of the other player and multiplied them by -1 in the greedy allocation conditions to compare the slopes over time for generous and greedy allocations. We then standardized the allocations and fairness ratings and

subjected them to a multivariate ANOVA with the between-subject factors 2 (valence: generous vs. greedy split) x 2 (time: recent vs. far past) x 2 (voluntary vs. involuntary). The hypothesized three-way interaction of valence x time x intentionality was significant ($F(1, 323) = 7.73$, $p = .006$).

For voluntary allocations to the other player, the interaction contrast was significant ($t(323) = 2.37$, $p = .018$). A previous voluntary, generous allocation by the other player impacted participants' allocations to the other player more when it had occurred recently ($M = 1.05$, $SD = 0.84$) than when it had occurred in the past ($M = 0.32$, $SD = 0.68$, contrast $t(323) = 4.41$, $p < .001$). In contrast, a previous voluntary, greedy allocation by the other player impacted participants' allocations equally, whether it had occurred recently ($M = 0.14$, $SD = 0.73$) or in the past ($M = -0.04$, $SD = 0.73$, contrast $t(323) = 1.08$, $p = .28$). No such interaction pattern was found for involuntary allocations by the other player (interaction contrast: $t(323) = 0.21$, $p = .84$; see figure 2A for raw means).

A similar pattern was observed for fairness ratings (interaction contrast: $t(323) = 2.39$, $p = .017$). A previous voluntary, generous allocation by the other player impacted participants' fairness ratings more when it had occurred recently ($M = 1.24$, $SD = 0.65$) than when it had occurred in the past ($M = 0.64$, $SD = 0.46$, contrast $t(323) = 4.83$, $p < .001$), but a previous voluntary, greedy allocation by the other player impacted participants' fairness ratings equally, whether it had occurred recently ($M = 0.14$, $SD = 0.40$) or in the past ($M = -0.04$, $SD = 0.57$, contrast $t(323) = 1.47$, $p = .14$). Again, no interaction pattern was found for involuntary allocations by the other player (interaction contrast: $t(323) = 0.05$, $p = .34$; see figure 2B for raw means).

Figure 2A: Allocation to other player as a function of previous generous/greedy and voluntary/involuntary splits by the other player in past or recent rounds in experiment 2. Error bars represent +/- 1 standard errors.

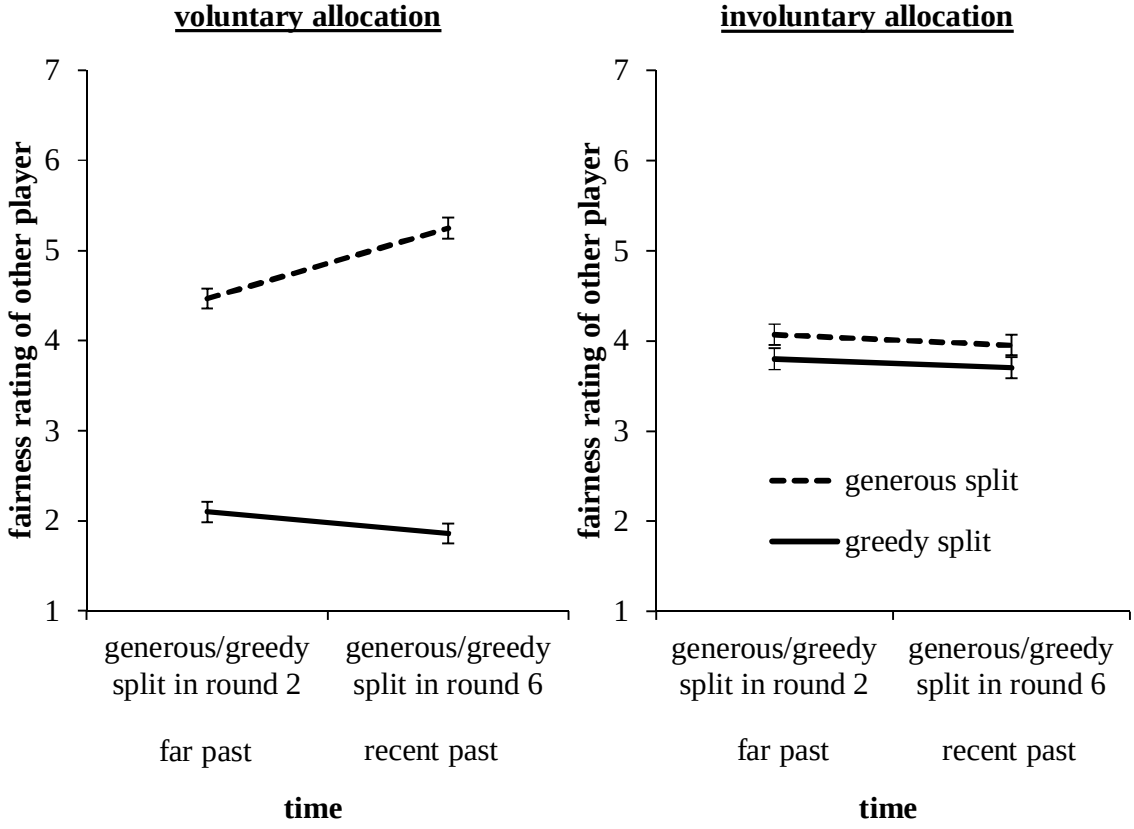


Discussion

The results of experiment 2 show that past generous behaviors were discounted more than past greedy behaviors. However, such differential discounting of past behaviors was only found when the other player had engaged in these activities voluntarily. Involuntary actions were not diagnostic of the other player's intentions, and had hence no impact on fairness ratings of, and allocation of money to the other player. This pattern of results provides further evidence that differential discounting is not solely due to the extremity of the behaviors or the greater attention-grabbing power of negative social information. More importantly, the results demonstrate that past behaviors not only affect overall impressions and fairness judgments, but

influence the behavior of others toward that person. Having been greedy once in a dictator game, even when it was long time ago, substantially decreases people's willingness to share money with that person, whereas the effect of having been generous quickly dissipates over time.

Figure 2B: Fairness ratings of other player as a function of previous generous/greedy and voluntary/involuntary splits by the other player in past or recent rounds in experiment 3. Error bars represent +/- 1 standard errors.



EXPERIMENT 3

Experiment 3 employed the same design and used the same material as Experiment 1. In addition to the positive/negative moral behavior, we added 4 conditions with positive/negative behaviors in the domain of intelligence. The experiment thus employed a 2 (valence: positive vs. negative behavior) x 2 (time: recent versus far past) x 2 (domain: morality vs. intelligence)

between-subjects design, with an additional control condition in both domains. We set N = 50 per cell since we collected data online and anticipated effects to be weaker.

Participants and Procedure

Five hundred ninety-seven participants were recruited through the Amazon Mechanical Turk for a study on impression formation and paid $.30 ($M_{duration\ of\ study}$ = 3.13 minutes, SD = 1.26). Ninety-seven participants (19.4%) failed to correctly answer the comprehension questions, their responses were excluded, leaving 500 responses for further analysis (229 males, M_{age} = 31.24, SD = 11.96).

The background information provided about the target person was the same as in experiment 1. The critical positive/negative behavior was either related to the person's honesty (reporting or not reporting a lost purse to the police containing $10,000 in cash as in experiment 1) or the person's intelligence (proving a very hard math theorem, or needing a calculator to add up two numbers of any kind; Skowronski & Carlston, 1987). Participants were asked to rate the target person's honesty/intelligence on 9-point Likert scales with the endpoints, 'extremely dishonest/stupid' (1) and 'extremely honest/intelligent' (9).

Results

Manipulation check. We conducted two (positive vs. neutral vs. negative behavior) one-way ANOVAs, separately for honesty and intelligence ratings. Both were significant (honesty: $F(2, 252) = 376.68$, $p < .001$; intelligence: $F(2, 242) = 180.96$, $p < .001$). For the average honesty/intelligence ratings and t-tests between the three conditions see table 3.

Table 3: Average morality and intelligence ratings of the target person as a function of her/his previous behaviors in experiment 3. Standard deviations are in parentheses.

previous behaviors by target person	morality ratings of target person	*t*-tests morality	intelligence ratings of target person	*t*-tests intelligence
1 positive behavior, all others neutral	8.29 (1.23)	$t(158) = 10.49$ $p < .001$	8.39 (0.77)	$t(145) = 11.06$ $p < .001$
only neutral behaviors	6.09 (1.29)	$t(146) = 7.16$ $p < .001$	6.79 (0.98)	$t(147) = 12.46$ $p < .001$
1 negative behavior, all others neutral	3.05 (1.50)		4.79 (1.82)	

Discounting of negative and positive behaviors. As in experiment 1, we mean-centered morality and intelligence ratings and multiplied them by -1 in the negative behavior conditions to compare the slopes over time for positive and negative behaviors. We ran an ANOVA with the between-subject factors 2 (valence: positive vs. negative) x 2 (time: recent vs. far past) x 2 (domain: ability vs. morality). A main effect for domain emerged ($F(1, 286.88) = 35.22$, $p < .001$), and more importantly, the hypothesized three-way interaction of valence x time x domain was significant ($F(1, 286.88) = 4.02$, $p = .046$; degrees of freedom and F- and t-values are adjusted for unequal variances).

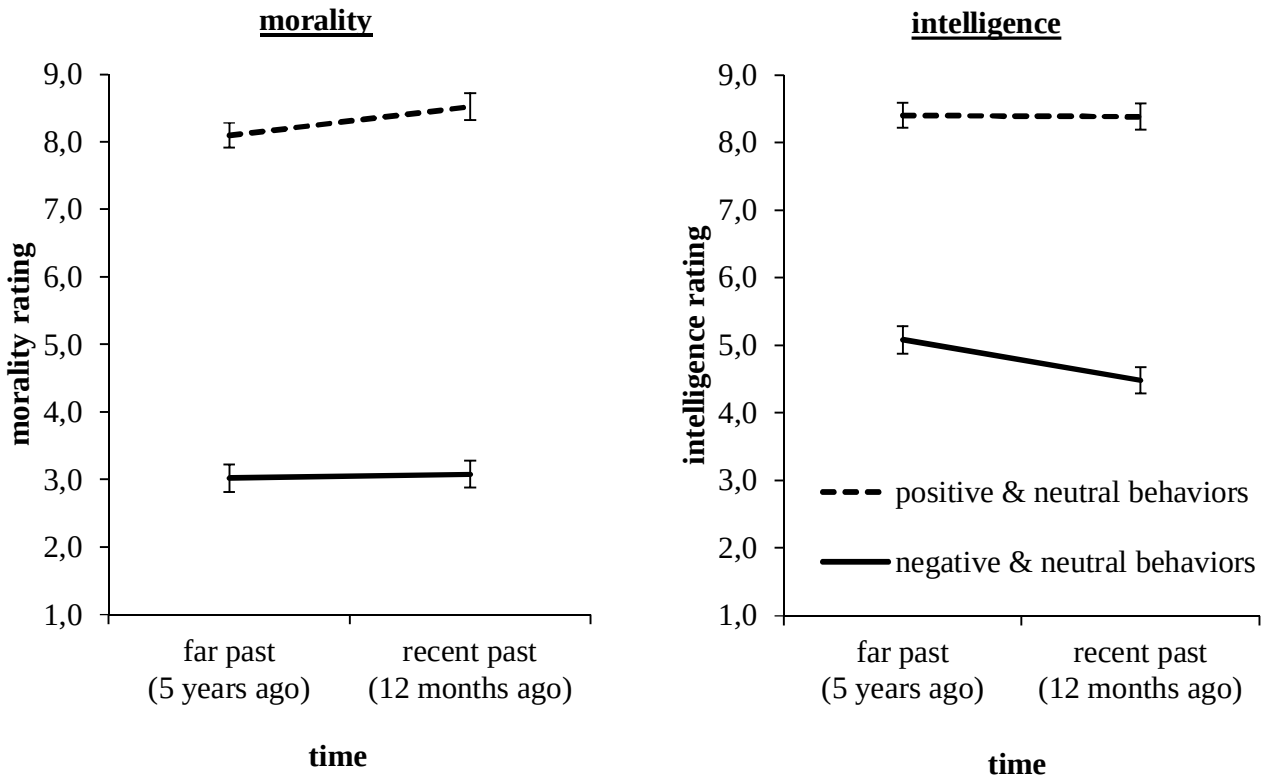
For morality ratings, the interaction contrast did not reach significance but was in the hypothesized direction ($t(161.48) = 1.27$, $p = .21$). As predicted, it made little difference whether the immoral behavior had occurred 5 years ago ($M = 2.79$, $SD = 1.51$) or 12 months ago ($M = 2.73$, $SD = 1.51$, contrast $t(92.61) = 0.19$, $p = .85$), but it made a bigger difference whether the moral behavior had occurred 5 years ago ($M = 2.27$, $SD = 1.56$) or 12 months ago ($M = 2.70$, $SD = 0.67$, contrast $t(74.98) = 1.89$, $p = .063$).

For intelligence ratings, the interaction contrast did again not reach significance but was in the hypothesized direction ($t(130.67) = 1.56$, $p = .12$). As predicted and opposite to the pattern of morality ratings, it made little difference whether the intelligent behavior had occurred 5 years ago ($M = 1.83$, $SD = 0.76$) or 12 months ago ($M = 1.80$, $SD = 0.80$, contrast $t(93.80) = 0.11$, $p = .91$), but it made a bigger difference whether the stupid behavior had occurred 5 years ago ($M = 1.50$, $SD = 1.82$) or 12 months ago ($M = 2.10$, $SD = 1.79$, contrast $t(95.94) = 1.89$, $p = .10$; see figure 3 for raw means).

Discussion

The same pattern of less discounting of immoral versus moral behaviors as in experiment 1 was found in experiment 3, although the interaction contrast failed to reach the conventional significance level of 5%. More importantly, though, the interaction pattern for intelligence ratings was opposite to that of morality ratings, here negative behaviors were discounted more than positive behaviors. This three-way interaction is predicted by the diagnosticity explanation, but cannot be explained by the three alternative accounts. All three alternative accounts, negative behaviors are more extreme, negative social information is more attention-grabbing (Pratto & John, 1991), and negative events elicit more causal attribution than positive events (Peeters & Czapinski, 1990), predict less discounting of negative behaviors in both domains, morality and intelligence judgments.

Figure 3: Morality and intelligence ratings of the target person as a function of behaviors displayed in the recent (12 months) or far past (5 years ago) in experiment 3. Error bars represent +/- 1 standard errors.



GENERAL DISCUSSION

Bad does not only have a stronger impact than good (Baumeister et al., 2001; Rozin & Royzman, 2001), its impact also fades slower over time than that of good. However, there is an important exception to this general principle. When good is more diagnostic – for example, in the domain of ability, positive behaviors are more diagnostic than negative behaviors – good not only has a stronger impact on impression formation (Skowronksi & Carlston, 1987, 1989), but – as we show – has also a longer lasting impact on impressions of, and behavior toward the person that displayed the behavior.

Chapter 4

Forming an impression about others based on disclosures – The unexpected effect of similar self-disclosures

with Alessandro Acquisti and Francesca Gino

Abstract

Intimate, embarrassing, even self-incriminating public disclosures have become quite common in social media, and they may constitute the starting point to form an impression about the person who disclosed. As social norms evolve and sensitive disclosures in Web 2.0 become more acceptable, the question arises as to whether such disclosures will still be considered diagnostic of an individual's personality, and will therefore affect impression formation, when most people may have their own embarrassing records online. In this paper we analyze the effect of sensitive disclosures on the impressions one will form of others who made similar disclosures. In three studies, we test the paradoxical effect that disclosure of a sensitive trait can have on the judgment of others with a similar pattern of disclosure. Both with observational data coming from voluntary disclosures (Study 1), and with experimental data where probability of disclosure is manipulated (Study 2), we find that people who disclose an embarrassing trait or behavior are more judgmental of others who made similar disclosures. Moreover, people who admit that they engaged in unethical behaviors are more judgmental of others who also made similar admissions (Study 3). We provide evidence that this is due to the fact that admitting to an unethical behavior reduces guilt and provides the individual license to be judgmental about others' unethical behaviors.

INTRODUCTION

Social network sites, such as Facebook or Twitter, video sharing sites, such as YouTube, and other social media, such as photo sharing sites or blogs, have become repositories of increasingly large amounts of personal information – sometimes inappropriate, sometimes even self-incriminating – that users, for a variety of different reasons, willingly post about themselves. At the most general level, disclosures constitute a means of connection with one's "friends:" we share personal information because we want to keep in touch with family and friends; we want to share our interests and ideas with similar others; we search for group support in order to achieve similar goals (boyd & Ellison, 2007). Social media have made this process particularly easy and effortless, and have thus incentivized information sharing.

Some features of information shared online make it fundamentally different from information shared offline. First of all, digital traces are potentially permanent: even if a piece of information is taken down from a website, and even though most companies have policies that allow them to store users' information for a limited amount of time, what has been publicly posted has been accessible to "friends" or (depending on the policies of the host website and the choices of the user) all Internet users for at least *some* time. This means that publicly posted content is essentially undeletable. Second, information posted online has the potential for vast outreach and is extremely easy to search, both in terms of time and costs – even more so with semantic search,[8] which seems to be the promising future of search engines, and with software for natural language interpretation.[9] Third, information posted online can be easily taken out of

[8] See for instance Google's recent announcement of their new semantic search tool, Knowledge Graph: http://googleblog.blogspot.co.uk/2012/05/introducing-knowledge-graph-things-not.html.
[9] See for instance Apple's Siri (Speech Interpretation and Recognition Interface): http://www.apple.com/iphone/features/siri.html.

context and misinterpreted – a feature not unique to online information, but that becomes especially critical in combination with the previous two.

One would expect that these characteristics would make online ways of expression more prudent as compared to their offline counterparts, as the effects of the former in terms of impression management and reputation (Gosling, Gazing, & Vazire, 2007; Lampe, Ellison, & Steinfeld, 2007) may involve much larger audiences for a much longer time. But what we observe on networking sites and other social media does not seem to meet this expectation.

It could be argued that, since in a few years from now most of us will have sensitive information about ourselves on the Internet, sometimes even not revealed by us directly, and the social norms about what is considered an appropriate online disclosure will have fundamentally changed, information that might be categorized today as embarrassing or harmful for one's reputation will not receive as much attention in the near future as it does now. This paper addresses the question of whether that is necessarily the case. Norms regarding online disclosures might change, but the effect that those disclosures have on the impressions that others will form may be more resilient to technological innovations.

Consider employers who, as part of the hiring process, need to screen job applicants. It used to be the case that a resume, and possibly references, represented the focal instrument for candidates' screening. With the successful proliferation of social media, though, plenty of *personal* information, voluntarily provided by users, can easily be gathered by employers about potential future employees, and used to form an overall impression of candidates beyond their strictly professional capabilities. In fact, according to a 2011 survey[10] by Reppler, a social media

[10] See http://www.blueoceanmi.com/assets/news/1339543867WFS-April2012.pdf, last accessed on November 20, 2012.

monitoring service, almost 70% of surveyed employers rejected a candidate at least once based on content found on a social networking site. Similar results were found independently by several other entities,[11] such as Microsoft,[12] confirming that social media background checking has been growing in the past few years, at times even approaching the border of legitimacy – e.g., when individuals are required to provide passwords to their social networking profiles, or to connect to their employers through the online network, a practice that has recently caught the attention of the US Senate.[13] Cases under investigation left aside, the search for publicly available information about job candidates is a common practice among employers, therefore personal information posted online contributes significantly to the impression others form about us, and is all but inconsequential.

The question of whether online disclosures will keep on affecting impression formation in the same way that they do today can only be answered through conjectures, while waiting for time to provide empirical evidence. In this paper, we address a question that is more limited in scope, but nonetheless informative, since it involves cognitive processes that are more stable than social norms, and less sensitive to external shocks like technological innovations. Specifically, we are interested in the effect of the disclosure of a certain trait or behavior on the impression that one will form of others who made similar disclosures. Is it the case that disclosure of a certain negative personal trait or behavior will make one more lenient towards others who made similar disclosures? Or could it be that disclosure of that trait or behavior actually represents a way of "coming out clean," so that one will feel licensed to be more

[11] See for instance Search Engine Journal (http://www.searchenginejournal.com/the-growth-of-social-media-an-infographic/32788/) and CareerBuilder (http://thenextweb.com/socialmedia/2012/04/18/survey-37-of-your-prospective-employers-are-looking-you-up-on-facebook/), last accessed on November 20, 2012.
[12] See go.microsoft.com/?linkid=9709510, last accessed on November 20, 2012.
[13] See http://www.zdnet.com/blog/facebook/us-senators-investigate-employers-asking-for-facebook-passwords/10834 for details.

judgmental towards others, with no worries about being regarded as hypocritical, or simply unable to look at one's own personal flaws?

Literature on similarity does not provide a conclusive answer about what should be the effect of similar disclosures on impression formation. On the one hand, several studies provide support for the hypothesis that similarity causes attraction (e.g. Berscheid, 1985; Byrne, 1971; Byrne, Clore, & Worchel, 1966; Newcomb, 1956; for a review, see Byrne, 1997), dissimilarity causes repulsion (Rosenbaum, 1986), and that group membership and expected similarity mediate this relationship (Chen & Kenrick, 2002). On the other hand, similarity was found to produce avoidance or dislike if the similar person was described as part of a stigmatized group, such as a former mental patient (Novak & Lerner, 1968), or an obnoxious individual (Taylor & Mettee, 1971). Other studies found that the case where "opposites attract" is more than just an exception (Aron, Steele, Kashdan, & Perez, 2006; Dryer & Horowitz, 1997; Grush, Clore, & Costin, 1975; Heider, 1958; O'Leary & Smith, 1991). Finally, some more recent studies have looked into Construal Level Theory to reconcile these contradictory results (Liviatan, Trope, & Liberman, 2008; McCarthy & Skowronski, 2011). These studies suggest that people will tend to focus on lower level (subordinate and secondary) information to form impressions of similar (thus psychologically close) others, while higher level (superordinate and primary) construals will play a central role in forming impressions of dissimilar (distant) others. Therefore, either attraction or contempt should result depending on the degree of similarity on high-level construal qualities, such as ideals or goals, as opposed to lower-level construal qualities, such as habits. This research though, does not test this prediction empirically, focusing on the mental representation of other people's actions or on personality inference based on a list of traits, and it

does not look at the specific effect of disclosure. Furthermore, in these studies a judgment is not expressed on the same trait, attitude, action or behavior that is used to induce similarity.

In this paper, we measure (rather than *induce*) similarity more subtly, and then ask for judgments on a target person based on the same traits that define similarity itself. Consider the case of a person whose Facebook profile hosts some personal embarrassing or compromising pictures, for example photos of this person being drunk. Based on the personal content published on social media, what impression will this person form of others whose profiles also contain similar pictures? The answer is not trivial. On the one hand, this person may form a positive impression about others who also posted sensitive personal material online because she feels similar to them. This person may also form no specific impression about others with similar online records because she may think that, in fact, most people have their own "skeletons on the Internet," so online profiles do not really provide enough (or at all relevant) information to form an impression. On the other hand, she could react to such information by applying double standards for herself and others. Consistent with observer bias in attribution theory (Jones & Nisbett, 1971) and with literature on moral hypocrisy (Lammers, Stapel, & Galinsky, 2010) and ethical inconsistencies (Ayal & Gino, 2011; Barkan, Ayal, Gino, & Ariely, 2012; Gino & Ariely, 2012), people may be able to easily justify their own compromising online information (not necessarily unethical), while they will be less likely to do so for information regarding others.[14] As a result, they may still be judgmental of others that made similar online disclosures. In fact, they may express even harsher judgments based on such disclosures as compared to others who never really disclosed anything embarrassing about themselves because, consistent with cognitive dissonance (Festinger, 1957; Cooper, 2007), this latter type of people will feel less

[14] This effect differs from moral licensing (e.g. Cain, Loewenstein, & Moore, 2005) as it refers to one's judgment of others' behavior, not to one's overindulging in a certain behavior because its disclosure gives license to do so.

need to distance themselves from the target person, and may thus express less extreme judgments. Finally, and specific to unethical behaviors, it could be the case that admitting to one's wrongdoings may reduce the feeling of guilt that arises from immoral actions, thus increasing moral hypocrisy (Polman & Ruttan, 2012) and providing a sense of lower rather than higher self-threat (as standard cognitive dissonance theory would predict), and the feeling that one has almost a *right* to be judgmental about others' unethical behaviors.

OVERVIEW OF THE EMPIRICAL STUDIES

The effect of disclosure on the judgment one will express of others who made similar disclosures depends on many factors that contribute to generate an intricate model of interactions. It is useful to distinguish some of these factors here in order to better explain the general design of our experiments and the differences across them. First of all, we analyze the effect of disclosure D by subject S on the judgment that S will express of a target person T. T may be presented as either *possessing* a certain trait, or as publicly *disclosing* that trait. In the former case, S will express a judgment based on the specific trait that T possesses; in the latter case, S will express a judgment of the public disclosure of that trait. This is an important distinction, because, while one may have no particular opinion about a certain trait or behavior by others, it could be the very choice of publicly displaying it that generates a negative impression. We would expect therefore different reactions to information that the target person T willingly decides to disclose, and information that regards T but that is disclosed by somebody else. Disclosure D could refer to a specific trait or behavior B, or to a general and comprehensive index of disclosure, and it could be a public, voluntary disclosure P or a private, elicited admission A. B could represent either an embarrassing trait or behavior E, or an unethical one U. Finally, the judgment we focus on could be an attitudinal (M_1) or a behavioral measure (M_2).

In Study 1, we use attitudinal measures M_1 (willingness to hire a target person T) to test whether public, voluntary disclosure P of specific embarrassing behaviors E correlates negatively with the judgment of others who made similar disclosures, and whether this effect vanishes when T is not personally responsible for the disclosure, which was made by someone else instead. A fading effect would suggest that, irrespective of disclosure P by the subjects, judgments about T are affected by her willful, public disclosure of a certain trait or behavior, rather than by the trait or behavior itself, which could be made public by someone else, possibly even without T being aware of it. To provide an explicit example of this phenomenon, consider an employer who finds on the social media profile of a job applicant some inappropriate photos − for instance, photos of the applicant being drunk. The employer may have nothing against the fact that the applicant likes alcohol, but even if she has her own embarrassing online records, she may, at the same time, have a negative opinion about aspiring employees who, in her eyes, chose to publicly present themselves as drunkards, or simply did not think about the consequences that this disclosure would have had on their online reputation. On the other hand, the employer's opinion may be less affected by those pictures if they were not published by the applicant herself, but by someone else, who might have had a variety of different reasons to post them, such as making fun of a certain person, or willingly embarrassing her, or simply not considering the repercussions of the publication of those photos. More formally, we can summarize our hypotheses as follows:

H1A: public disclosure of an embarrassing behavior is negatively correlated with attitudinal measures of impression formation regarding a target person who made similar disclosures.

H1B: impression formation regarding a target person is influenced by that person's self-disclosures of embarrassing behaviors, and not merely by the embarrassing behaviors themselves.

Study 1 was an observational study that realistically mimicked everyday life scenarios, and had the major advantage of using voluntary (as opposed to elicited) disclosures. This feature was at the same time a strength and a limitation of this study: since people self-select into a disclosing or non-disclosing type, deciding whether or not to disclose certain information about their personal traits or behaviors, Study 1 does not allow for any causal inference about the effect of disclosure itself.

In order to solve this self-selection issue, in Study 2 we manipulated the probability of disclosure experimentally, thus addressing causality more directly. In order to do so, though, we had to give up the dimension of public disclosure, and used elicited private disclosure A instead. Specifically, by private disclosure we mean admissions that the subject makes only to the researchers as part of the study. This study employed the same attitudinal measure M_1 used in Study 1 to test whether specific embarrassing private disclosures E negatively affect the impression one will form of a target person T who also made similar disclosures. We can formally summarize our hypothesis as follows:

H2: private disclosure of an embarrassing behavior decreases attitudinal measures of impression formation regarding a target person who made similar disclosures.

Finally, Studies 3A and 3B focused on specific unethical behaviors U and used behavioral measures M_2 (money allocations in a dictator game) to test whether people who *admit* to having behaved unethically are more judgmental of others who made similar admissions, and

whether people who *behave* unethically, but are not given a chance to admit it, are more judgmental of others who also behaved unethically as compared to people who did not behave unethically. More formally, Studies 3A and 3B, respectively, test the following hypotheses:

H3A: admissions of unethical behaviors decrease behavioral measures of impression formation regarding a target person who also made similar admissions.

H3B: absent a chance of admission, behaving unethically decreases behavioral measures of impression formation regarding a target person who also behaved unethically.

While the element of disclosure makes the first hypothesis an entirely new contribution to the literature on unethical behaviors and impression formation, the latter issue was partially addressed by Barkan et al. (2012), but our study differs from theirs in that 1) it is not based on recall of generic unethical behaviors the subject is asked to write about, but rather on actual immoral actions performed by the subject while participating in the study; and 2) it uses behavioral rather than attitudinal measures.

All our studies were in the form of online surveys or games, and subjects were recruited from the Amazon Mechanical Turk, an online platform that allows quick recruitment of a vast and demographically diverse pool of subjects.

STUDY 1

Study 1 was a two-phase experiment: we used the first phase to identify some personal information that subjects had publicly disclosed in the past, and in the second phase, we asked subjects to express a judgment about someone else who had also made a similar disclosure.

In phase 1, we recruited 350 subjects (44% female, average age = 30.6, SD = 10.9) to participate in a 10-minute study on "Online Sharing," and paid them $0.50. Subjects were asked if they had ever posted specific pieces of content online (on social networks such as Facebook or LinkedIn, blogs or other publicly accessible Internet websites). The content we asked about included items commonly found in Web 2.0 applications, such as demographic information or organization of events; embarrassing or sensitive items, such as salacious or compromising photos, information regarding personal health or finances, political and religious views; content directly related to work environment, such as complaint about one's job or colleagues; offensive comments, such as swearwords or other offenses directed to other people (see the screenshot in the Appendix for the complete list of questions asked in this phase). We also asked whether subjects ever removed any specific content from the website where it was posted, and if so, why they removed it. The purpose of this last question was to try and understand the mental process behind online disclosures and consequent judgment of others. Just because one disclosed a particular trait or behavior online, it doesn't mean that one believes it was a wise thing to do. While, as we argued in the introduction to this chapter, "un-doing" is typically impossible when it comes to online publication of sensitive content, one could still reduce this dissonance by removing posted content from public exposure.

Four weeks after data collection for phase 1 was finished (a time lapse that assured our subjects considered the two phases as separate studies, and thus limited demand effects), we posted phase 2 of the study on the Amazon Mechanical Turk, and made it visible only to workers who had previously participated in phase 1. Phase 2 consisted in expressing a judgment about a person based on personal information about her retrieved online – more specifically, the focal piece of information we used was a drunken photo (a piece of content we asked about in Phase

1). The reasons why we picked this particular scenario were that 1) we had enough variation in the responses to the question on posting drunken photos online in phase 1, with approximately 30% of the respondents responding affirmatively; and 2) content about drinking is one of the items that, according to the Reppler survey mentioned above, determine rejection of job applicants, so it makes the study realistic and the results applicable to real world scenarios.

In phase 2, subjects were recruited for a 10-minute survey on "Hiring Decisions," and they were paid $0.50. Since we did not specifically invite all subjects from phase 1 to participate in a follow-up (to avoid that they made a connection between the two phases, and tried to recall how they previously answered to questions in phase 1), but simply made the study visible on the Mechanical Turk only to them for two weeks, we obtained responses only from 153 subjects (40% female, average age = 30.1, SD = 10.3), or about 44% of all the subjects who took part in phase 1. They were instructed to imagine that they worked for the Human Resources department of an advertisement company, and that they were in the process of hiring a new employee (target person T) who was described as qualified for the job – a similar scenario to the one used by Barkan et al. (2012). The hiring practices of the company required them to search for publicly available information about T in social media, and the result of this search included basic demographic information, the network of people T was connected to, and photos of T being drunk, either published by T personally (Personal condition) or by a friend of T (Other condition). Subjects were randomized to one of these two conditions. This setup allowed us to distinguish between judgments led by the behavior at stake (drinking alcohol) or rather by the *voluntary public disclosure* of that behavior.

After reading about T, subjects were asked how likely they would be to hire T on a scale from 1 (Very Unlikely) to 7 (Very Likely). This measure of impression formation was our main

dependent variable of interest. Subjects were also asked whether in their opinion T showed lack of prudence and whether she showed bad judgment on a scale from 1 (Strongly disagree) to 7 (Strongly agree). We used the responses to these questions as manipulation checks, to make sure subjects paid attention to the type of disclosure at the basis of the impression formation process – specifically, the fact that in the Personal condition, T had personally disclosed embarrassing or compromising information about herself, while in the Other condition, she had engaged in a certain behavior but she had not chosen to publicly disclose it, and others disclosed it instead. Finally, in order to control for psychological distance, which is a main determinant of impression formation (McCarthy & Skowronski, 2011), participants answered 8 questions regarding how close they felt to T (relatedness sub-scale from the Intrinsic Motivation Inventory; Ryan, Koestner, & Deci, 1991).

Results

Manipulation checks. Seventy-six (44% female, average age = 30.6, SD = 10.2) and 77 subjects (36% female, average age = 29.7, SD = 10.5) were randomized to the Personal and the Other condition, respectively. Subjects considered T as a person showing higher imprudence and worse judgment in the Personal condition ($M_{imprudence}$ = 4.79, SD = 1.37; $M_{bad\_judgment}$ = 4.99, SD = 1.50) than in the Other condition ($M_{imprudence}$ = 3.60, SD = 1.50, $t(151) = 5.14$, $p < .0001$; $M_{bad\_judgment}$ = 3.67, SD = 1.63; $t(151) = 5.17$, $p < .0001$). Evaluations regarding lack of prudence and bad judgments were not significantly affected by whether subjects had posted drunken photos of themselves (p-values larger than .10 in both conditions).

Impression Formation. In order to test H1A, or whether public disclosure of an embarrassing behavior correlates negatively with the impression one will form of others who

also publicly disclosed that behavior, and H1B, or whether impression formation would not be affected by embarrassing disclosures if they are made by someone other than the target person, we estimated the following basic model for each of the 2 conditions:

$$y_i = \alpha + \beta * Self\_disclosure_i + \gamma * Relatedness_i + \delta * Demographics_i + \varepsilon_i \quad (1)$$

where *y* is the main dependent variable of interest (willingness to hire T), a categorical variable coded as "low" for the negative side of the scale (Very Unlikely/Unlikely/Somewhat Unlikely to hire), "medium" for the mid value of the scale (Undecided), and "high" for the positive side of the scale. *Self_disclosure* is a binary variable equal to 1 if the subject said, in phase 1 of the study, that she had posted online a drunken photo of herself, and represents our main explanatory variable; *Relatedness* is an index of relatedness obtained as the average of the 8 items composing the corresponding scale; *Demographics* is a vector including gender (binary variable equal to 1 if the subject was male) and age; and ε is a random error. Since *y* is an ordered categorical variable, we estimated an ordered logit model. A Brant test could not reject the null that the parallel regression assumption held (chi-square p-values larger than 0.7 both for each explanatory variable separately and for the full model).

Tables 1A and 1B summarize the results (marginal effects) of the estimation of the model for the Personal and the Other condition, respectively.[15]

[15] When we estimate the model by pulling the data from the two conditions together, and add the condition and its interaction with *Self_disclosure* as regressors, the interaction is not significant, suggesting that the difference between the coefficients in the two conditions, estimated separately, is not significant.

Table 1A

Willingness to hire	***Unlikely to Hire***	***Undecided***	***Likely to Hire***
Self_disclosure	.332*** (.132)	-.193* (.097)	-.139*** (.059)
Relatedness	-.405*** (.090)	.200** (.091)	.205*** (.050)
Male	-.246** (.130)	.124 (.081)	.122** (.064)
Age	-.005 (.007)	.003 (.003)	.003 (.003)
	N = 76 Chi-squared (4) = 23.73		

Marginal effects of ordered logit model (1) for Personal condition − Study 1. Standard errors in brackets. *** indicates significance at the 1% level; ** at 5% level; * at 10% level.

Table 1B

Willingness to hire	***Unlikely to Hire***	***Undecided***	***Likely to Hire***
Self_disclosure	-.018 (.059)	-.025 (.086)	.043 (.145)
Relatedness	-.133*** (.040)	-.177*** (.073)	.310*** (.091)
Male	.064** (.051)	.091 (.072)	-.155 (.119)
Age	-.001 (.003)	-.002 (.003)	.003 (.007)
	N = 77 Chi-squared (4) = 13.11		

Marginal effects of ordered logit model (1) for Other condition − Study 1. Standard errors in brackets. *** indicates significance at the 1% level; ** at 5% level.

These results support hypotheses H1A and H1B: From the negative coefficient on *Self_disclosure* in the High category in Table 1A, we can infer that, in the case where the target person had posted embarrassing material about herself online, subjects were less likely to hire her (less likely to categorize the target person as one they would be very likely, likely, or somewhat likely to hire) if they reported that they had also posted embarrassing material about themselves online. Since none of the coefficients on *Self_disclosure* is significant in Table 1B, we can infer that this effect vanishes completely in the case where the target person did not post embarrassing material about herself, but someone else did so instead. The raw mean of the likelihood to hire the target person was significantly higher in the Other condition (M_{hire} = 4.66, SD = 1.34) than in the Personal condition (M_{hire} = 3.64, SD = 1.50, t(151) = 4.416, p < .001). Not surprisingly, the degree to which the subject related to the target person (as measured by the coefficient on *Relatedness*) was always significant.

Finally, out of the 41 subjects who reported having posted a drunken photo of themselves (20 in the Personal and 21 in the Other condition), 18 (44%) also said that they had later on removed it, suggesting that, in hindsight, they may have deemed the posting inappropriate, and thus took action to reduce the resulting cognitive dissonance by somehow "un-doing" the publication of embarrassing material.[16]

In order to better understand the mental process behind this effect, and test whether it was due to cognitive dissonance, we also estimated two models similar to (1) for the Personal condition, but replacing *Self_disclosure* with either (Specification A) the Removal of the posted material (a binary variable equal to 1 if the subject said that she had removed the drunken photo

[16] Although we did not perform proper semantic or sentiment analysis, open ended responses confirmed that the reason why people removed the drunken photo from the online profile was mainly the realization that its public display could have had negative consequences for reputation, especially with colleagues or employers.

from her online profiles), or (Specification B) the interaction of *Self_disclosure* and *Removal* of the posted material, in order to obtain an estimate of the effect of a regretted self-disclosure on impression formation as compared to either non-regretted disclosure or non-regretted disclosure and non-disclosure, respectively. Tables A.1 and A.2 in the Appendix show the results of these estimations. For Specification A (Table A.1), we had very few observations, because only 20 subjects in the Personal condition reported that they removed the drunken photo from their profile. Still, the marginal effect on the lowest category of the likelihood to hire was positive and significant, suggesting that, conditional on having posted a drunken photo of themselves, subjects who removed it were less likely to hire the target person than subjects who never removed the photo. For Specification B (Table A.2), where we pool together subjects who never posted a drunken photo of themselves and subjects who posted it but did not regret it, we observe a significant effect on all three categories of the likelihood to hire. Specifically, can see that subjects who regretted posting a drunken photo of themselves were significantly less likely to hire the target person, as compared to subjects who either never posted that content or posted it and never removed it (and thus did not experience dissonance).

Although willingness to disclose was not randomized, and therefore subjects self-selected into a more open or a closer type, we can conclude that, based on embarrassing self-disclosures, such as a drunken photo, they formed negative impressions of others even though they had regretted disclosures themselves. In fact, we find that disclosure of embarrassing content is negatively correlated with attitudinal measures of impression formation about a target person who made similar disclosures.

STUDY 2

Study 2 aimed at establishing a causal effect of self-disclosures by randomizing participants to either a higher or a lower disclosure condition, and by measuring their impression of others who made similar disclosures. The scenario was similar to the one we used in Study 1, but it consisted of one unique phase. Subjects were recruited to take a 10-minute "Survey about behaviors and attitudes," and they were paid $0.50. After providing demographic information (gender and age), they were randomized to one of two conditions: either High or Low Disclosure. In order to elicit high or low willingness to disclose, we used a manipulation inspired by the findings in Frey (1986) and Singer, Hippler & Schwarz (1992), according to which anonymity and confidentiality reassurances can decrease willingness to answer personal questions, because they make privacy issues salient. Subjects in the Low Disclosure condition read the following notice:

"**IMPORTANT - PLEASE READ**

Information concerning the **confidentiality and anonymity** of your responses.

Please be advised that **maintaining the confidentiality and anonymity of your responses is of the utmost importance** to us. The following procedure will be used to guarantee confidentiality in analysis, publication, and presentation of any results.

Any sensitive information collected through this survey will be securely stored and password-protected. In addition, your survey responses will only be analyzed in aggregate, and any published results will only show aggregate data."

The initial warning (IMPORTANT - PLEASE READ) had larger, bold, red font, so that subjects would be more likely to notice it and read it carefully. In order to reinforce the effect even

further, we also asked subjects to explain how their responses would be treated using their own words. Subjects randomized to the High Disclosure condition saw no such notice, and were directly sent to the next part of the survey, where they were asked whether or not they had ever engaged in a variety of sensitive, embarrassing or unethical behaviors, such as smoking marijuana, being prejudiced, lying about income or education, badmouthing an employer (see the screenshot in the Appendix for a list of all 13 questions asked in this study). This way we elicited disclosure of sensitive behaviors, but willingness to disclose was experimentally manipulated.

The next part of the study consisted of a filler task. Subjects were shown easily recognizable drawings of four different kinds of activities (studying, eating, playing piano and playing tennis) in random order, and they were asked to describe those activities in one or two words (see a screenshot of one of the drawings in the Appendix). Finally, subjects were presented with the same hiring scenario used in phase 2 of Study 1, but this time the relevant online information about the job applicant consisted in pictures of the candidate smoking cannabis. Similarly to online content about the candidate consuming alcohol, not surprisingly content about the use of drugs was also, according to the Reppler survey mentioned before, one of the potential reasons for rejection of a job candidate based on the information retrieved on her/his online profile.

In order to test H2, we estimated model (1) using the Wald estimator, or instrumenting our possibly endogenous variable *Self_disclosure* with a dummy for the random assignment to one of the two conditions, High or Low Disclosure. The reason for using the randomized condition as an instrument in this setup is that, as in cases of imperfect compliance (Angrist, 1990; Angrist, Imbens & Rubin, 1996), it is impossible to experimentally *make* subjects disclose something: the best one can do is to make them more likely to do so. Thus, simply estimating the

effect of being in the High Disclosure condition doesn't really tell us the effect of disclosure, unless we divide it by the effect of being in the High Disclosure condition on the likelihood to disclose. This is precisely what the instrumental variable approach does. Allowing for heterogeneous effects of our manipulation, the β coefficient will represent the effect of disclosure on "compliers," or those who disclosed because they were randomly assigned to the High Disclosure condition, but would not have otherwise been willing to disclose. The model

$$y_i = \alpha + \beta * HighDisclosure_i + \gamma * Relatedness_i + \delta * Demographics_i + \varepsilon_i \quad (2)$$

has two interpretations. First, it is the reduced form for the instrumental variable regression above. Second, the coefficient on *HighDisclosure* can be interpreted as the "intention to disclose" (borrowing the terminology from Angrist, 1990), or the average effect of assigning someone to a treatment (low reassurances) that makes them more likely to disclose.

Results

Manipulation check. One-hundred-one subjects took part in Study 2 (46% female, average age = 28.8, SD = 10.2), 49 and 52 in the Low and High Disclosure condition, respectively. Overall willingness to disclose, as measured by the average willingness to admit an embarrassing or unethical behavior, was higher in the High than in the Low Disclosure condition ($M_{high} = 0.54$, $M_{low} = 0.45$; $t(99) = 2.28$, $p < .05$) for almost all questions, although the difference in admission rates was not always significant (see Table A.3 in the Appendix for a summary of admission rates for each individual question).

Impression formation. Table 2A and 2B summarize the results (marginal effects) of model (1), where *Self_disclosure* was instrumented with the dummy for High Disclosure, and model (2), respectively. For the estimation of these models, *y* was categorized as a binary

variable equal to 0 if the subject was Very Unlikely/Unlikely/Somewhat Unlikely to hire the target person or Undecided, and equal to 1 otherwise.[17]

The data provide empirical support for H2. The negative coefficient on *Self_disclosure* in Table 2A suggests that subjects who, due to our manipulation of willingness to disclose (low confidentiality reassurances), admitted to having used drugs were 54% less likely to hire a job applicant whose online profile contained information about drug use, than subjects who did not admit it. Moreover, from Table 2B, we can see that the average effect of the manipulation on willingness to hire the job candidate is negative and significant: subjects in the High Disclosure condition were about 18% less likely to express a favorable opinion about hiring the target person than those in the Low Disclosure condition.

Table 2A

Willingness to hire	
Self_disclosure	-.54*** (.049)
Relatedness	224*** (.041)
Male	.154** (.073)
Age	003 (.004)
	N = 101 Wald Chi-squared (8) = 84.89

Marginal effects of Instrumental Variable probit model (1) − Study 2. Standard errors in brackets. *** indicates significance at the 1% level; ** at 5% level.

[17] We tried to use STATA's conditional mixed process package (Roodman, 2009) for the estimation of an IV ordered probit model, but the algorithm did not converge with three categories for the dependent variable, possibly due to the very low number of "Undecided" responses (19 out of 101). We therefore coded *y* as a binary variable.

Table 2B

Willingness to hire	
HighDisclosure	-.181*** (.077)
Relatedness	.219*** (.044)
Male	.140** (.062)
Age	.003 (.003)
	N = 101 Wald Chi-squared (8) = 3.92

Marginal effects of probit model (2) − Study 2. Standard errors in brackets. *** indicates significance at the 1% level; ** indicates significance at the 5% level.

STUDY 3A

Study 1 and Study 2 showed that, consistent with cognitive dissonance, disclosure of embarrassing or compromising personal information makes one more judgmental about similar personal disclosures made by others. In Study 3A, we used behavioral (real money allocations in a dictator game) rather than attitudinal measures, we focused exclusively on unethical behaviors, and tested whether the mechanism of dissonance applies to this case as well. We hypothesized that, specifically for immoral actions, the process that leads to negative judgments about others could actually be different from ethical dissonance (Barkan et al., 2012).

We propose that disclosure could represent a means to reconcile one's wrongful actions with one's internal ethicality, or a way to set oneself free from the sense of guilt or shame caused by immoral actions. Once those actions are committed, the subject could either admit them, if given the opportunity, or deny them, thus lying and effectively behaving unethically twice. The

first alternative could be represented in the subject's mind as a way to redeem herself from the previous misdeed, thus resulting in lower self-reported guilt, rather in than in a feeling of self-threat (which is what ethical dissonance would predict). This mechanism would therefore make the subject feel free to express harsh judgments about the unethicality of other people, without worrying about appearing hypocritical or unable to recognize her own faults. A second lie, on the other hand, which would further threaten the subject's ethicality after an immoral action has been committed, would not provide the subject with that clean slate which would license a strict ethical judgment, and, we predict, would thus result in less negative reactions towards others' unethical behaviors. *Admissions* to one's moral transgressions (rather than transgressions alone) are therefore key to determining the response to other people's wrongdoings.

In order to test this theory, we designed a study where subjects had a chance to behave unethically (cheat in a Die-Throwing Game). In order to evaluate the effect of admissions, similarly to Study 2, we then randomized subjects into two conditions: in the Low Admission condition, they received a treatment that made them less likely to admit that they cheated; and in the High Admission condition, the treatment made them more likely to admit. The treatment consisted in the same mechanism used in Acquisti et al. (2012): admissions were elicited by informing how many subjects had (allegedly) admitted to cheating until then. This information was visually represented by colored histograms where the percentage of admissions was either very high (close to 75%) or very low (close to 15%; see Figure A.4 for the histogram shown in the High Admission condition). Just like in Study 2, we used the randomized condition as an instrument for the subject's admission to having behaved unethically, and estimated the average effect of admissions on money allocations in a dictator game played with another (fictitious) player who also made similar admissions.

More specifically, Study 3A was composed of several parts. In the first part, subjects played an online Die-Throwing Game (what Jiang, 2012 calls the Mind Game) where the outcome of each throw (15 rounds in total) determined their earnings. In this game, subjects were asked to decide which side of the die, the one facing up or the one facing down, would determine their earnings before each throw. Then they threw the die and, finally, reported which side they had chosen in their mind before the throw. This game gives players a chance to cheat by declaring a "lucky side" (the one that gives higher payoff regardless of the outcome of the throw), and it allows for an interesting within-subjects natural variation of the extent of the cheating. Because in a standard die the opposing sides' outcomes sum up to 7 (e.g., if the side facing up is 1, the one facing down is 6; if the side facing up is 3, the one facing down is 4), subjects can cheat by a lot if they take advantage of the lucky side (the one facing down) when the outcome of the throw is 1 or 2, or they can cheat by a little if they report the lucky side when the outcome of the throw is 3. We cannot know, of course, whether subjects cheated or not, but, if we needed it, we could estimate the extent of the cheating by comparing the earnings subjects obtained with the ones that chance would grant. Also, for our purposes, the important thing is not so much whether the subject cheated or not, but whether our disclosure manipulation made them more likely to admit it.

After playing the Die-Throwing Game for 15 rounds, the second part of the study began and subjects were randomized to one of the two conditions mentioned above. In the High (Low) Admission condition, they read the following message:

“After running this study several times, we feel that some of the participants, after seeing the outcome of a die throw in a given round, may have reported a "lucky side" rather than the actual chosen side, i.e. they may have reported the opposite side to the one they had actually chosen. For this reason, after reassuring them that their bonus would not be affected by their answer, we started to ask participants to tell us whether they had ever misreported their chosen side - and most of them told us they did at least once (the never did).

Here is a summary of how participants responded to the question of whether or not they had misreported their chosen side:”

Then subjects saw the histograms reporting either a vast majority of admissions or of non-admissions. This manipulation was meant to simulate the effect of social norms regarding disclosure, and to push subjects to be more or less likely to admit to whether they had cheated or not. After answering a comprehension question regarding the histograms, that besides constituting an attention check, also reinforced the manipulation, subjects were asked whether in any of the 15 rounds they played, they ever reported a different side than the one they had actually chosen. The possible answers were: Never, Once or Twice, Occasionally, Often, Always, which we later recoded as 0 for Never and 1 for the other choices. Admissions constituted our main explanatory variable of interest.

In the third part of the study, subjects answered one question on whether or not they perceived the die to be fair and two questions about probability (these questions served as filler tasks; the only question that may have been of partial interest, as an aside to our main research question, could have been the one about perceived fairness of the die, but the distribution of perceived fairness did not differ across conditions). Right after this filler task, subjects were presented with the 20-item version of the PANAS (Positive and Negative Affect Scale; Watson, Clark & Tellegen, 1988). Our measure of interest was only the feeling of guilt, but in order to

minimize demand effects, especially after an admission of cheating, we included all 20 items in random order.

Finally, the last part of the study consisted in a dictator game (available sum to be split: 10cents), which we called "Allocation Game" and which subjects played (allegedly) with another player, described as one who responded to the admission question with "Often," so s/he admitted that s/he often cheated.[18] Our dependent variable of interest was the amount of money that subjects decided to keep for themselves (or equivalently, the sum they left on the table for the other player). Then the final set of questions was the relatedness scale also used in Study 2. At the end of this dictator game, subjects were reminded of their earned final bonus (the sum of what they earned in the Die-Throwing Game and the amount they allocated to themselves as dictators in the Allocation Game), debriefed and paid.

In order to test H3A we estimated models (1) and (2), where *y* was the amount of money in cents that subjects kept for themselves in the dictator game, and *Self_disclosure* was a binary variable equal to 1 if the subject admitted to having cheated at least once in the Die-Throwing Game. Similar to Study 2, we instrumented *Self_disclosure* with the randomized condition, so as to obtain the same kind of Wald estimator (also in this experiment, the instrument was a dummy variable equal to 1 for the High Admission condition).

Results

Manipulation checks. Two-hundred subjects (44% female, average age = 28.3, SD = 9.2), 100 in each condition, were recruited to participate in a 15-minute study involving "Online games." They were paid $0.10 for participating and told that they could earn up to a $1 bonus,

[18] This game was developed very similarly to the one used in Chapter 3, except for the history of rounds played by the Recipient, which in this game was substituted by the information regarding the Recipient's admissions.

for a total of $1.10. Cents earned in the Die-Throwing game did not differ across conditions (Low Admission: M_{earned} = 62.6, SD = 9.4; High Admission: M_{earned} = 62.1, SD = 9.8, t(198) = .434, p > .10), suggesting that propensity to cheat was similar in the two conditions. As far as our manipulation of the probability of admitting to cheating is concerned, subjects were more likely to admit to their unethical behavior in the High ($M_{confess}$ = 0.63, SD = 0.48) than in the Low Admission condition ($M_{confess}$ = 0.44, SD = 0.50, chi-squared(1) = 7.255, p < .01).

Money allocations. Subjects kept a higher share of the money in the dictator game if they were in the High (M_{self} = 7.56, SD = 2.42) than in the Low Admission condition (M_{self} = 6.73, SD = 2.51, t(198) = -2.378, p < 0.05), suggesting that admissions of cheating did have an effect on the way they split the money with another confessed cheater.

In order to test H3A more formally, we estimated model (1) through instrumental variables − Table 3A summarizes the results of the estimation. Supporting our hypothesis, subjects who admitted to having cheated were harsher on average than those who did not, in the sense that they allocated more money to themselves (on average approximately 4 more cents) and less to the other player who also admitted cheating, as compared to subjects who did not admit it.

Table 3B summarizes the results of the estimation of model (2), which is the reduced form of model (1). We can interpret the coefficient on the dummy for High Admission as the average effect of the treatment on money allocation choices: subjects in the High Admission condition kept on average almost 1 cent more in the dictator game than subjects in the Low Admission condition.

Table 3A

Allocations to self	
Self_disclosure	4.191** (1.913)
Relatedness	-.514** (.247)
Male	.041 (.369)
Age	.039 (.032)
	N = 200 Chi-squared (4) = 5.82

Instrumental variable regression for model (1) − Study 3A. Standard errors in brackets. **indicates significance at the 5% level.

Table 3B

Allocations to self	
HighAdmission	.813** (.354)
Relatedness	-.051 (.167)
Male	.115 (.377)
Age	-.001 (.019)
	N = 101 F(4,195) = 1.46

Marginal effects of model (2) − Study 3A. Standard errors in brackets. ** indicates significance at the 5% level.

Regarding the test of our theory for the process explaining this behavior, subjects' average self-reported feeling of guilt (on a scale from 1, or Not at All, to 5, or Extremely; nobody used this final point of the scale, so we do not report it in our table below) did not vary significantly across conditions, although, as our conjecture would predict, the average was slightly higher in the Low (M_{guilt} = 1.49, SD = .63) than in the High Admission condition (M_{guilt} = 1.43, SD = .07). We also estimated an ordered probit model where the dependent variable was the self-reported feeling of guilt, and the admission was the endogenous regressor instrumented with the randomized condition (we added demographics just to be consistent with the other models, but these variables were usually not significant). Table 3C reports the results of our estimation.

Table 3C

Guilt	***Not at all***	***A little***	***Moderately***	***Quite a bit***
Self_disclosure	.405***	-.145***	-.201***	-.058
	(.089)	(.028)	(.065)	(.052)
Male	-.111	.047	.052	.011
	(.065)	(.032)	(.031)	(.010)
Age	.007	-.003	-.003	.001
	(.003)	(.001)	(.002)	(.001)
	N = 200 LR Chi-squared(6) = 23.87			

Marginal effects of IV ordered probit model (1) for *Guilt*, where *Self_disclosure* is instrumented with the condition − Study 3A. Standard errors in brackets. *** indicates significance at the 1% level.

We can see that the effect of admissions is significant for all categories except the highest one (Quite a bit), and that this effect was positive for the lowest category and negative for the

higher categories, confirming our conjecture that subjects who admitted to cheating felt less guilty than those who did not admit it. Although the reduced form of this model (an ordered probit model where the regressors were only the randomized condition and the demographics) was not significant, the effect was in the hypothesized direction.

STUDY 3B

Study 3B was similar to 3A but it allowed us to test a slightly different hypothesis (H3B). While not at the center of this paper, as it does not focus on the effects of disclosure, this hypothesis is still interesting, as it gets to the relationship between immoral behaviors and judgments of similar immoral behaviors.

In this study, subjects played an online Die-Throwing Game at which they could either cheat (Cheating condition) or not (No Cheating condition). In the latter condition, they threw a standard, fair, six-sided die for 15 times, and they earned money at each throw. The amount of money earned, in cents, corresponded to the number that resulted from the throw. For example, if the throw resulted in a 3, the subject earned 3 cents for that throw. Since we recorded the outcome of the die, we automatically summed the earned cents for the subject after every throw, so there was no possibility to game the system and earn more money than chance would allow. In the Cheating condition, subjects played the same game as in Study 3A.

After the 15 rounds of the Die-Throwing game, all subjects answered the same questions asked in the third part of Study 3A (probability questions and PANAS) and played the same Allocation Game.

Results

Manipulation checks. Two-hundred subjects (44% female, average age = 28.3, SD = 7.97), 102 in the No Cheating and 98 in the Cheating condition, were recruited and paid with the same strategy used in Study 3A. Cents earned in the Die-Throwing game were significantly lower in the No Cheating (M_{earned} = 52.3, SD = 6.0) than in the Cheating condition (M_{earned} = 61.7, SD = 9.6, t(198) = -8.303, $p < 0.0001$). This suggests that, as expected, at least some people cheated in the latter condition.

Money allocations. Contrary to what H3B would predict, subjects in the No Cheating condition allocated significantly more money to themselves (M_{self} = 7.42, SD = 2.71) than subjects in the Cheating condition (M_{self} = 6.70, SD = 2.20, t(198) = 2.049, $p < .05$). This speaks against H3B, because it indicates that subjects who did not behave unethically were on average harsher towards cheaters than subjects who had an opportunity to behave unethically (and, as suggested by the earnings statistics, it is quite likely that on average they did). However, subjects in the No Cheating condition may have also realized that, as opposed to the player they were paired with for the Allocation Game, they could not take advantage of the opportunity to inflate their bonuses, and therefore they may simply have tried to get the most out of the last game they could play. This design does not allow us to distinguish between these (and possibly other) explanations for this finding, so we leave the test of H3B to future research.[19]

[19] One simple modification to our design that could address this confound could be the type of game subjects play in the last part of the study. Instead of a dictator game, we could use an ultimatum game where subjects would play second, so they could punish the confessed cheater (by rejecting their offer) in a way that is also costly for them.

CONCLUSIONS

Web 2.0 technologies, such as social media, have become increasingly popular in the last decade. The number of users and the amount of content shared through these websites has grown exponentially. Consider this simple statistic: it was estimated that in 2000, 85 billion physical photos were shot worldwide[20] − in 2011, according to Facebook engineers on the network's Photo application, 6 billion photos per month were uploaded by Facebook users alone.[21] Because, on the one hand, these technologies have lowered the time- and effort-related costs of sharing, and because, on the other hand, people are subject to biases such as hyperbolic time discounting and immediate gratification (Acquisti &Grossklags, 2004, 2005) when it comes to considering future costs associated with information sharing activities, it is no wonder that people have started sharing more and more.

Besides this merely quantitative growth, content shared has also increased in terms of variety (from thoughts, to photos, to videos) and scope: people tend to share all kinds of personal information on social media, from medical to financial information, from religious and political views to professional information. Lately, we have been assisting to the rise of sharing services that allow people to, either in an anonymous (examples include: grouphug.us, now a Facebook-based page; anonymousconfessions.com; postsecret.com) or − quite strikingly − identified fashion, publish embarrassing, sensitive, or at times even self-incriminating or compromising information. Notorious is the case of Hunter Moore's porn site IsAnyoneUp (see http://www.thedaily.com/article/2012/08/23/082312-news-hunter-moore/ for news about his comeback, dating back to August 2012), but other less infamous examples include people's

[20] See http://blog.1000memories.com/94-number-of-photos-ever-taken-digital-and-analog-in-shoebox, last retrieved on December 2, 2012.
[21] http://www.quora.com/How-many-photos-are-uploaded-to-Facebook-each-day/all_comments/Justin-Mitchell, , last retrieved on December 2, 2012.

voluntary posts to Facebook or Twitter, which sometimes cost users their job or their reputation, leading to regrets (Wang et al., 2011).

It seems as though social norms about online disclosures are changing fast, with more and more sensitive material being circulated on the Internet. Like in the story of the boy who cried the wolf, does this mean that soon enough we will stop paying attention to this information altogether? Will we stop considering this information as diagnostic of the kind of person that posted it, because everybody has similar online records? Will we still consider it inappropriate, or compromising as time passes and a larger number people share it? More to the research question addressed in this paper, does the fact that people post sensitive material about themselves online imply that soon they will be de-sensitized about other people's similar material? In three studies, we provide evidence of the opposite result: Mistakes made regarding hasty online publications, or even admissions to certain sensitive behaviors, do not necessarily make people more lenient in their judgments towards others making similar mistakes. In fact, our results suggest that people become stricter about embarrassing disclosures by others.

Study 1 looked at the correlation between propensity to post embarrassing material online and impressions about others who have done similar disclosures. The study finds that people who self-reported that their online records included embarrassing material, such as drunken photos, would be less likely to hire a qualified job candidate who had drunken photos on her social media profile than people who never posted such material in the first place. This negative impression about the target person is only formed when she personally posted the compromising photo online − perhaps a reassuring result from the perspective of social media users, considering that the Internet and Web 2.0 technologies have vastly reduced the control that people have over the publication of their personal information by others.

Study 2 went beyond the correlational analysis of Study 1 and was characterized by an experimental manipulation of the propensity to disclose, in order to test whether disclosure of embarrassing information had a causal effect on impression formation. Using the same scenario-based attitudinal measure of impression formation (willingness to hire a qualified candidate) as in Study 1, Study 2 had a design with two conditions, where subjects were either more or less reassured about the anonymity and confidentiality of the responses provided. Higher salience of these issues in the condition with strong reassurances made them indeed less willing to admit certain sensitive, embarrassing or unethical behaviors. This experimental manipulation was then used in the analysis as an instrument for the admission itself. The results of Study 2 suggest that people who admitted to having tried drugs were less likely to hire a job candidate who had posted material related to her drug use on her social media profile, as compared to people who did not admit it. We estimate that the Local Average Treatment Effect, as it was named by Angrist, Imbens and Rubin (1996), was as high as 54%: subjects who, due to our experimental manipulation, admitted that they had done use of drugs were 54% less likely to hire a qualified job candidate who had posted online a photo of herself using drugs as compared to subjects who did not admit it.

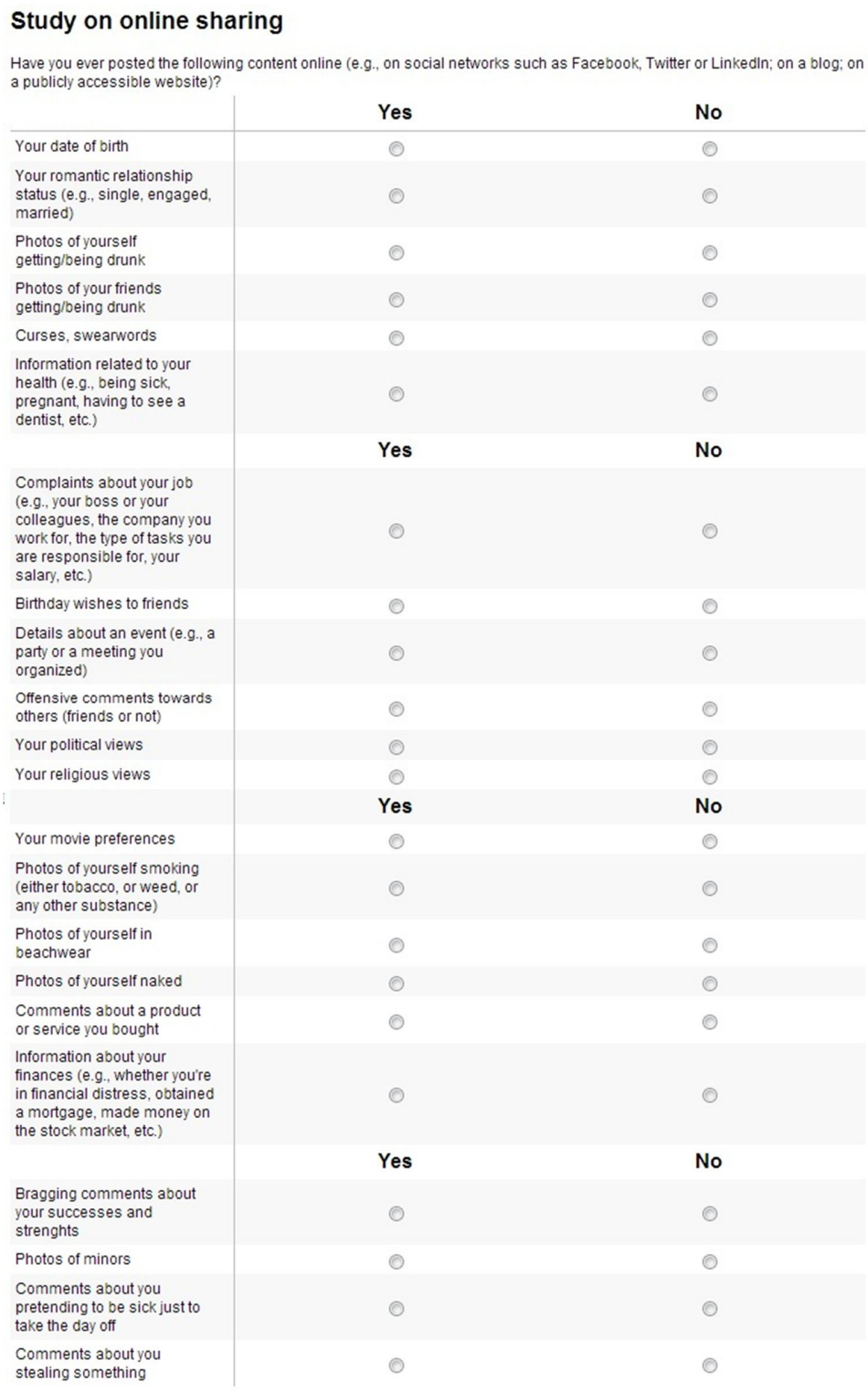
Finally, Study 3 focused on disclosure of unethical behaviors, and used a behavioral measure to test the hypothesis that admissions can actually make one more judgmental about others who made similar admissions. The results indicate that, in a dictator game, people tend to allocate less money to a confessed game cheater if they admitted to having cheated themselves than if they didn't admit it. This result might either seem counter-intuitive, or just a simple example of moral hypocrisy, but we propose a new mechanism that could explain it. Admissions could reduce the sense of guilt that arises after an immoral action, such as cheating in a game, is

committed. The feeling of redemption that admissions provide can thus bring people to express harsher judgments about the unethical behaviors of other people without risking to feel and appear hypocritical, or unable to recognize their own ethical flaws. Contrary to standard dissonance arguments, whereby it is the threat to the self that determines a negative reaction towards other people's immoral actions, we suggest that it is the sense of relief and self-enhancement arising from admissions of one's own misdeeds that provides people with the license to be judgmental about others. Our results, although not significant, provide some qualitative empirical support for this mechanism.

Overall, our studies provide empirical evidence for the hypothesis that even though more and more people may have embarrassing or compromising online records, not necessarily this will make them more lenient in their judgments about others with similar records. In fact, online disclosures could make them even more judgmental about others' disclosures. Social norms about online disclosures may be changing quickly, but the way people form impressions about others based on those disclosures may be much less responsive.

APPENDIX

Figure A.1: Questions in Study 1 – Phase 1.



Study on online sharing

Have you ever posted the following content online (e.g., on social networks such as Facebook, Twitter or LinkedIn; on a blog; on a publicly accessible website)?

	Yes	No
Your date of birth	○	○
Your romantic relationship status (e.g., single, engaged, married)	○	○
Photos of yourself getting/being drunk	○	○
Photos of your friends getting/being drunk	○	○
Curses, swearwords	○	○
Information related to your health (e.g., being sick, pregnant, having to see a dentist, etc.)	○	○
	Yes	**No**
Complaints about your job (e.g., your boss or your colleagues, the company you work for, the type of tasks you are responsible for, your salary, etc.)	○	○
Birthday wishes to friends	○	○
Details about an event (e.g., a party or a meeting you organized)	○	○
Offensive comments towards others (friends or not)	○	○
Your political views	○	○
Your religious views	○	○
	Yes	**No**
Your movie preferences	○	○
Photos of yourself smoking (either tobacco, or weed, or any other substance)	○	○
Photos of yourself in beachwear	○	○
Photos of yourself naked	○	○
Comments about a product or service you bought	○	○
Information about your finances (e.g., whether you're in financial distress, obtained a mortgage, made money on the stock market, etc.)	○	○
	Yes	**No**
Bragging comments about your successes and strenghts	○	○
Photos of minors	○	○
Comments about you pretending to be sick just to take the day off	○	○
Comments about you stealing something	○	○

Figure A.2: Questions in Study 2.

Study on behaviors and attitudes

Here is a list of various kinds of behaviors and situations. Did you ever engage in these behaviors, or did you ever find yourself in these situations?

	Yes	No
Have you ever called in sick when you were not sick?	○	○
Have you ever taken credit for someone else's work?	○	○
Have you ever smoked marijuana (i.e., pot, weed)?	○	○
Have you ever claimed to have education you didn't actually obtain?	○	○
Have you ever made up a serious excuse, such as grave illness or death in the family, to get out of doing something?	○	○
Have you ever tried to gain access to someone else's email account (e.g., a partner's, friend's, colleague's) without their knowledge or consent?	○	○
Have you ever downloaded pirated material (e.g., songs, videos, software) from the Internet?	○	○
Do you always turn the lights out at home and work, even if you're feeling lazy?	○	○
Have you ever claimed to have higher skills at a certain task than you actually had (e.g., claiming to be an advanced user of a certain tool, software or technology when you only had basic knowledge of it; claiming to be an advanced player of a certain sport or game when you were only a beginner)?	○	○
Have you ever revealed information that was supposed to be kept private, such as a secret or confidential information?	○	○
Are you always unbiased and unprejudiced in your interactions with people?	○	○
Have you ever lied about your income/wealth or that of your family?	○	○
Have you ever badmouthed any of your employers or work superiors?	○	○

Figure A.3: Filler task in Study 2.

Study on behaviors and attitudes

For this part of the study, we ask you to please describe in one or two words the behaviors and activities you see in the 4 images on this page.

What is this person doing?

Figure A.4: Histogram shown in the High Admission condition in Study 3.

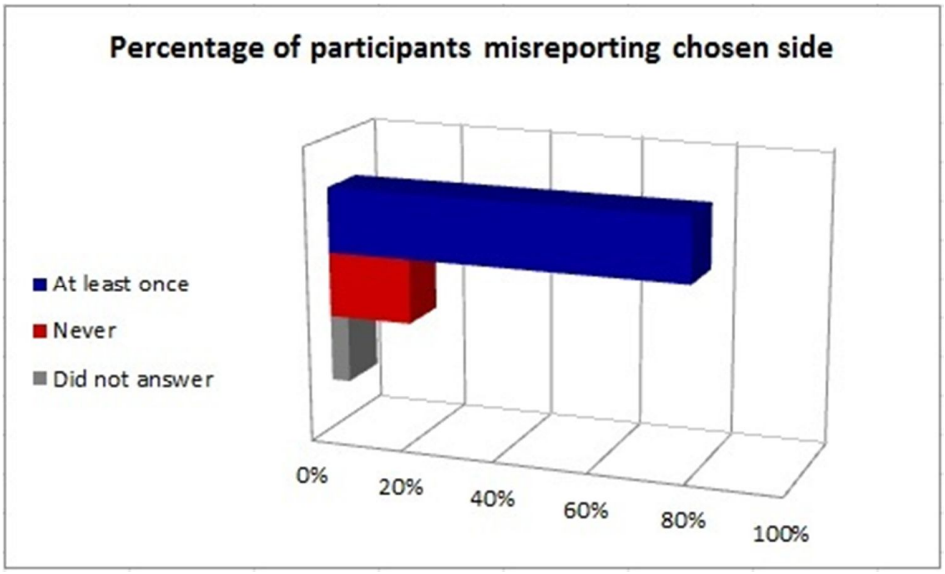


Table A.1

Willingness to hire	***Unlikely to Hire***	***Undecided***	***Unlikely to Hire***
Removal	.585** (.287)	-.459 (.317)	-.125 (.161)
Relatedness	-.943* (.510)	.728 (.701)	.215 (.229)
Male	-.130 (.240)	.105 (.191)	.025 (.064)
Age	-.036 (.043)	.028 (.043)	.008 (.009)
	N = 20 Chi-squared (4) = 13.82		

Marginal effects of modified ordered logit model (1) (replacing *self-disclosure* with *removal* of the content conditional on *self-disclosure* being equal to 1) for Personal condition − Study 1. Standard error in brackets. ** indicates significance at the 5% level; * at 10% level.

Table A.2

Willingness to hire	***Unlikely to Hire***	***Undecided***	***Unlikely to Hire***
Self_disc*Removal	.388*** (.127)	-.253*** (.010)	-.135*** (.053)
Relatedness	-.384*** (.080)	.183** (.082)	.201*** (.051)
Male	-.197 (.139)	.096 (.080)	.101 (.068)
Age	-.007 (.007)	.003 (.003)	.003 (.003)
	N = 76 Chi-squared (4) = 28.20		

Marginal effects of modified ordered logit model (1) (replacing *self-disclosure* with the interaction of *self-disclosure* and *removal* of the content) for Personal condition − Study 1.
Standard error in brackets. *** indicates significance at the 1% level; ** at 5% level.

Table A.3

Item	Percentage of affirmative admissions	
	Low Discl	**High Discl**
1. Have you ever called in sick when you were not sick?*	55%	73%
2. Have you ever taken credit for someone else's work?*	28%	40%
3. Have you ever smoked marijuana (i.e. pot, weed)?*	51%	67%
4. Have you ever claimed to have education you didn't actually obtain?**	18%	25%
5. Have you ever made up a serious excuse, such as grave illness or death in the family, to get out of doing something?	28%	27%
6. Have you ever tried to gain access to someone else's email account (e.g., a partner's, friend's, colleague's) without their knowledge or consent?*	37%	48%
7. Have you ever downloaded pirated material (e.g., songs, videos, software) from the Internet?	65%	77%
8. Do you always turn the lights out at home and work, even if you're feeling lazy?	49%	54%
9. Have you ever claimed to have higher skills at a certain task than you actually had (e.g., claiming to be an advanced user of a certain tool, software or technology when you only had basic knowledge of it; claiming to be an advanced player of a certain sport or game when you were only a beginner)?*	47%	61%
10. Have you ever revealed information that was supposed to be kept private, such as a secret or confidential information?	59%	60%
11. Are you always unbiased and unprejudiced in your interactions with people?	63%	75%
12. Have you ever lied about your income/wealth or that of your family?	24%	31%
13. Have you ever badmouthed any of your employers or work superiors?*	59%	67%

Admission rates by question and condition (listed in order of presentation) − Study 2. Questions 8 and 11 coded as affirmative admissions if the subject responded 'No.' *chi square test significant at $p \leq 0.05$ (2-sided), **chi square test significant at $p \leq 0.01$ (2-sided), using Bonferroni correction.

References

Acquisti, A. (2004). "Privacy in electronic commerce and the economics of immediate gratification." Electronic Commerce Conference, New York.

Acquisti, A. and Gross, R. (2006). "Imagined communities: awareness, information sharing, and privacy on the Facebook." *Proceedings of the Sixth Workshop on Privacy Enhancing Technologies*.

Acquisti, A. and Grossklags, J. (2004). "Losses, gains, and hyperbolic discounting: Privacy attitudes and privacy behavior." In J. Camp and R. Lewis (Eds.), "The Economics of Information Security," Kluwer, pp. 179-186.

Acquisti, A. and Grossklags, J. (2005). "Privacy and rationality in individual decision making." *IEEE Security and Privacy*, 3(1): 26-33.

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

Acquisti, A. and Varian, H. (2005). "Conditioning prices on purchase history." *Marketing Science*, 24(3): 367-381.

Angrist, J. (1990). "Lifetime earnings and the Vietnam era draft lottery: Evidence from Social Security administrative records." *American Economic Review*, 80: 313-335.

Angrist, J., Imbens, G. and Rubin, D. (1996). "Identification of causal effects using instrumental variables." *Journal of the American Statistical Association*, 91: 444-55.

Anton, A. I., Earp, J. B., and Young, J. D. (2010). "How Internet users' privacy concerns have evolved since 2002." *IEEE Security & Privacy*, 8(1): 21-27.

Aron, A., Steele, J. L., Kashdan, T. B., and Perez, M. (2006). "When similars do not attract: Tests of a prediction from the self-expansion model." *Personal Relationships*, 13: 387-396.

Ayal, S. and Gino, F. (2011). "Honest rationales for dishonest behavior." In M. Mikulincer & P. R. Shaver (Eds.), "The social psychology of morality: Exploring the causes of good and evil." Washington, DC: American Psychological Association.

Bargh, J. A., Chen, M., and Burrows, L. (1996). "Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action." *Journal of Personality and Social Psychology*, 71: 230-44.

Bargh, J. A., McKenna, K. Y. A., and Fitzsimon, G. M. (2002). "Can you see the real me? Activation and expression of the 'true self' on the Internet." *Journal of Social Issues*, 58(1): 33-48.

Barkan, R., Ayal, S., Gino, F., and Ariely, D. (2012). "The pot calling the kettle black: Contrast response to ethical dissonance." *Journal of Experimental Psychology: General*, 141 (4): 757-773.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D. (2001). "Bad is stronger than good." *Review of General Psychology*, 5: 323-70.

Berscheid, E. (1985). "Interpersonal attraction." In G. Lindzey & E. Aronson (Eds.), "Handbook of social psychology." New York: Random House.

Boekaerts, M., Pintrich, P. R., and Zeidner, M. (Eds.) (2000). “Handbook of self-regulation.” Elsevier, Amsterdam, The Netherlands.

boyd, d. (2004). “Friendster and publicly articulated social networks.” *Proceedings of ACM Conference on Human Factors in Computing Systems* (pp. 1279–1282). New York: ACM Press.

boyd, d. and Ellison, N. B. (2007). “Social network sites: Definition, history, and scholarship.” *Journal of Computer-Mediated Communication*, 13(1): 210-230.

Brickman, P., Coates, D., and Janoff-Bulman, R. (1978). “Lottery winners and accident victims: Is happiness relative?” *Journal of Personality and Social Psychology*, 36: 917-27.

Broadbent, D. E. (1957). “A mechanical model for human attention and immediate memory.” *Psychological Review*, 64(3): 205-215.

Broadbent, D. E. (1982). “Task combination and selective intake of information.” *Acta Psychologica*, 50(3): 253-290.

Brown, J. D. (1986). “Evaluations of self and others: Self-enhancement biases in social judgments.” *Social Cognition*, 4(4): 353-376.

Brown, J. S. and Duguid, P. (2000). “The Social Life of Information.” Harvard Business Press, Cambridge, MA.

Byrne, D. (1971). “The attraction paradigm.” New York: Academic Press.

Byrne, D. (1997). “An overview (and underview) of research and theory within the attraction paradigm.” *Journal of Social and Personal Relationships*, 14: 417-431.

Byrne, D., Clore, G. L., and Worchel, P. (1966). "Effect of economic similarity-dissimilarity on interpersonal attraction." *Journal of Personality and Social Psychology*, 4: 220-224.

Cahill, C., Llewelyn, S. P., and Pearson, C. (1991). "Long-term effects of sexual abuse which occurred in childhood: A review." *British Journal of Clinical Psychology*, 30, 117-130.

Cain, D. M., Loewenstein, G., and Moore, D. A. (2005). "The dirt on coming clean: Perverse effects of disclosing conflicts of interest." *Journal of Legal Studies*, 34: 1-25.

Chen, F. F. and Kenrick, D. T. (2002). "Repulsion or attraction? Group membership and assumed attitude similarity." Journal of Personality and Social Psychology, 83(1): 111-125.

CNEWS, Canada, Jan. 17th 2007, URL: http://cnews.canoe.ca/CNEWS/Canada/2007/01/17/3394584-sun.html, accessed on 4/13/2012.

ConsumersUnion.org (2008). "Americans extremely concerned about Internet privacy." Available at http://www.consumersunion.org/pub/core_telecom_and_utilities/006189.html.

Cooper, J. (2007). "Cognitive Dissonance: 50 Years of a Classic Theory." Sage, London.

Crawford, V. P. (2003). "Lying for strategic advantage: Rational and boundedly rational misrepresentation of intentions." *American Economic Review*, 93: 133-149.

Culnan, M. J. (1993). "How did they get my name? An exploratory investigation of consumer attitudes toward secondary information use." *MIS Quarterly*, 17(3): 341-363.

De Angelis, M., Bonezzi, A., Peluso, A. M., Rucker, D. D., and Costabile, M. (forthcoming). "On braggarts and gossips: A self-enhancement account of word-of-mouth generation and transmission." *Journal of Marketing Research*.

Dryer, C. D. and Horowitz, L. M. (1997). "When do opposites attract? Interpersonal complementarity versus similarity." *Journal of Personality and Social Psychology*, 72: 592-603.

Dukas, R. (2004). "Causes and consequences of limited attention." *Brain, Behavior and Evolution*, 63: 197-210.

Elgesem, D. (1996). "Privacy, respect for persons, and risk." In C. Ess (Ed.), "Philosophical Perspectives on Computer-Mediated Communication." New York, State University of New York Press.

Epley, N., Keysar, B., Van Boven, L., and Gilovich, T. (2004). "Perspective taking as egocentric adjustment." *Journal of Personality and Social Psychology*, 87: 327-339.

Federal Trade Commission (2000). "Privacy online: Fair information practices in the electronic marketplace." Available at http://www.ftc.gov/reports/privacy2000/privacy2000.pdf.

Festinger, L. (1957). "A Theory of Cognitive Dissonance." Stanford Univ. Press, Stanford, CA.

Fleming, J., Mullen, P. E., Sibthorpe, B., and Bammer, G. (1999). "The long-term impact of childhood sexual abuse in Australian women." *Child Abuse and Neglect*, 23: 145-159.

Floridi, L. (2007). "A look into the future impact of ICT on our lives." *The Information Society*, 23(1), 59-64.

Floridi, L. (2010). “Information. A very short introduction.” Oxford: Oxford University Press.

Forsythe, R., Horowitz, J. L., Savin, N. E., and Sefton, M. (1994). “Fairness in simple bargaining experiments.” *Games and Economic Behavior*, 6: 347-369.

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002). “Time discounting and time preference: A critical review.” *Journal of Economic Literature*, 40: 351-401.

Frey, J. H. (1986). “An experiment with a confidentiality reminder in a telephone survey.” *Public Opinion Quarterly*, 50:267-69.

Fried, C. (1984). “Privacy.” In F.D. Schoeman (Ed.), “Philosophical dimensions of privacy.” New York, Cambridge University Press.

Galinsky, A. D. (2002). “Creating and reducing intergroup conflict: The role of perspective-taking in affecting out-group evaluations.” In H. Sondak (Ed.), “Research on managing groups and teams: toward phenomenology of groups and group membership.” 4: 85-113. Greenwich, CT: Elsevier/JAI.

Galinsky, A. D., Magee, J. C., Inesi, M. E., and Gruenfeld, D. H. (2006). “Power and perspectives not taken.” *Psychological Science*, 17(12): 1068-1074.

Gilovich, T., Medvec, V.H., and Savitsky, K. (2000). “The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one’s own actions and appearance.” *Journal of Personality and Social Psychology*, 78: 211-222.

Gino, F. and Ariely, D. (2012). “The dark side of creativity: original thinkers can be more dishonest.” *Journal of Personality and Social Psychology*, 102(3): 445-459.

Gosling, S., Gaddis, S. and Vazire, S. (2007). "Personality impressions based on Facebook profiles." *Proceedings of the International Conference on Weblogs and Social Media* (ICWSM), Boulder, CO.

Grush, J. E., Clore, G. L., and Costin, F. (1975). "Dissimilarity and attraction: When difference makes a difference." *Journal of Personality and Social Psychology*, 32(5): 783-789.

Harris, P. (1996). "Sufficient grounds for optimism? The relationship between perceived controllability and optimistic bias." *Journal of Social and Clinical Psychology*, 15: 9-52.

Harris Interactive (2002). "First major post-9-11 privacy survey finds consumers demanding companies do more to protect privacy; public wants company privacy policies to be independently verified." See: http://prn.to/KQdZwT.

Hayek, F. A. (1945). "The use of knowledge in society." *American Economic Review*, 35(4): 519-530.

Heider, F. (1958). "The psychology of interpersonal relations." Oxford: Wiley.

Hilbert, M. and Lopez, P. (2011). "The world's technological capacity to store, communicate and compute information." *Science*, 332: 60-65.

Hindujaa, S. and Patchin, J. W. (2008). "Personal information of adolescents on the Internet: A quantitative content analysis of MySpace." *Journal of Adolescence*, 31(1): 125-146.

Ho, T.-H., Camerer, C., and Weigelt, K. (1998). "Iterated dominance and iterated best response in experimental p-beauty contests." *American Economic Review*, 88: 947-969.

Hoadley, C. M., Xu, H., Lee, J. J., and Rosson, M. B. (2010). "Privacy as information access and illusory control: The case of the Facebook News Feed privacy outcry." *Electronic Commerce Research and Applications*, 9(1):50-60.

Hughes, T. P. (1994). "Technological momentum." In M. R. Smith and L. Marx (Eds.), "Does technology drive history," Vol. 101, pp. 101-113). MIT Press.

Jiang, T. (2012). "The mind game: Invisible cheating and inferable intentions." *Katholieke Universiteit Leuven Discussion Paper*, 309.

John, L., Acquisti, A., and Loewenstein, G. (2011). "The best of strangers: Context-dependent willingness to divulge sensitive information."*Journal of Consumer Research*, 37(5): 858-873.

Johnston, W. A. and Dark, V. (1986). "Selective attention." *Annual Review of Psychology*, 37: 43-75.

Joinson, A.N., Woodley, A., and Reips, U.D. (2007). "Personalization, authentication and self-disclosure in self-administered Internet surveys." *Computers in Human Behavior*, 23:275-285.

Jones, E. E., and Nisbett, R. E. (1971). "The actor and the observer: Divergent perceptions of the causes of behavior." Morristown, NJ: General Learning Press.

Keller Fay Group (2006). "Single-Source WOM Measurement." Available at http://www.kellerfay.com/news/WOMMA%20-%20KF%20Paper%2011-2006.pdf. Last accessed on June 2, 2012.

Kahneman, D. (1973). "Attention and effort." New York, Prentice Hall.

Kang, J. (1998). “Information privacy in cyberspace transactions.” *Stanford Law Review*, 50:1193-1294.

Klamma, R., Cao, Y., and Spaniol, M. (2007). “Watching the blogosphere: Knowledge sharing in the Web 2.0.” *International Conference on Weblogs and Social Media*, Boulder, Colorado, USA.

Klein, W. M. and Kunda, Z. (1994). “Exaggerated self-assessments and the preference for controllable risks.” *Organizational Behavior and Human Decision Processes*, 59: 410-427.

Kolar, D. W., Funder, D. C., and Colvin, C. R. (1996). “Comparing the accuracy of personality judgments by the self and knowledgeable others.” *Journal of Personality*, 64(2): 311-337.

Krueger, J. (1998). “Enhancement bias in description of self and others.” *Personality and Social Psychology Bulletin*, 24(5): 505-516.

Lachter, J., Forster, K. I., and Ruthruff, E. (2004). “Forty-five years after Broadbent (1958): Still no identification without attention.” *Psychological Review*, 111(4): 880-913.

Lammers, J., Stapel, D. A., and Galinsky, A. D. (2010). “Power increases hypocrisy: Moralizing in reasoning, immorality in behavior.” *Psychological Science*, 21: 737-744.

Lampe, C., Ellison, N. and Steinfield, C. (2007). “A familiar Face(book): profile elements as signals in an online social network.” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, San Jose, CA.

Laufer, R.S. and Wolfe, R. (1977). “Privacy as a concept and a social issue: A multidimensional developmental theory.” *Journal of Social Issues*, 33:22-41.

Lessig, L. (2002). “Privacy as property.” *Social Research*, 69: 247-269.

Litan, R. E. and Rivlin, A. M. (2001). “Projecting the economic impact of the Internet.” *American Economic Review*, 91(2): 313-317.

Liviatan, I., Trope, Y., and Liberman, N. (2008). “Interpersonal similarity as a social distance dimension: Implications for perceptions of others’ actions.” *Journal of Experimental Social Psychology*, 44: 1256-1269.

Loewenstein, G., Weber, E. U., Hsee, C. K., and Welch, N. (2001). “Risk as feelings.” *Psychological Bulletin*, 127(2): 267-286.

MacKinnon, D. P., Fairchild, A. J., and Fritz, M. S. (2007). “Mediation analysis.” *Annual Review of Psychology*, 58:593 -614.

McCarthy, R. J. and Skowronski, J. J. (2011). “You’re getting warmer: Level of construal affects the impact of central traits on impression formation.” *Journal of Experimental Social Psychology*, 47: 1304-1307.

McDonald, A. and Cranor, L. F. (2008). “The Cost of Reading Privacy Policies.” *I/S: A Journal of Law and Policy for the Information Society*. Privacy Year in Review issue.

McDonald, A. and Cranor, L. F. (2010). “Americans’ attitudes about internet behavioral advertising practices.” *Proceedings of the 9th annual ACM workshop on Privacy in the electronic society*, pp. 63-72. ACM, New York, NY.

Miller, A. R. (1971). “The assault on privacy.” Ann Arbor, The University of Michigan.

Nagel, R. (1995). “Unraveling in guessing games: An experimental study.” *American Economic Review*, 85: 1313-1326.

Neisser, U. (1967). “Cognitive psychology.” New York, Appleton.

Newcomb, T. M. (1956). “The prediction of interpersonal attraction.” *American Psychologist*, 11: 575-586.

Nissenbaum, H. (2004). “Privacy as contextual integrity.” *Washington Law Review*, 79: 119-158.

Novak, D. W. and Lerner, M. J. (1968). “Rejection as a consequence of perceived similarity.” *Journal of Personality and Social Psychology*, 9(2): 147-152.

Nordgren, L. F., Van der Pligt, J., and Van Harreveld, F. (2007). “Unpacking perceived control in risk perception: The mediating role of anticipated regret.” *Journal of Behavioral Decision Making*, 20: 533-544.

O’Leary, K. D., and Smith, D. A. (1991). “Marital interaction.” *Annual Review of Psychology*, 42: 191-212.

Pashler, H. E. (1998). “The psychology of attention.” MIT Press, Cambridge.

Peeters, G. and Czapinski, J. (1990). “Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects.” *European Review of Social Psychology*, 1: 33-60.

Peltzman, S. (1975). “The effects of automobile safety regulation.” *Journal of Political Economy*, 83: 677-726.

Petronio, S. S. (2002). “Boundaries of privacy: dialectics of disclosure.” SUNY Press.

Phelps, J., Nowak, G., and Ferrell, E. (2000). “Privacy concerns and consumer willingness to provide personal information.” *Journal of Public Policy & Marketing*, 19(1):27-41.

Polman, E. and Ruttan, R. L. (2012). “Effects of Anger, Guilt, and Envy on Moral Hypocrisy.” Personality and Social Psychology Bulletin, 38: 129-139.

Pratto, F. and John, O. P. (1991). “Automatic vigilance: The attention-grabbing power of negative social information.” *Journal of Personality and Social Psychology*, 61: 380-91.

Preacher, K. J. and Hayes, A. F. (2004). “SPSS and SAS procedures for estimating indirect effects in simple mediation models.” *Behavior Research Methods, Instruments, and Computers*, 36:717-731.

Preacher, K. J. and Hayes, A. F. (2008). “Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models.” *Behavior Research Methods*, 40:879-891.

Rafaeli, S. and Raban, D. R. (2005). “Information sharing online: A research challenge.” International Journal of Knowledge and Learning, 1(1&2): 62-79.

Reeder, G. D. and Brewer, M. B. (1979). “A schematic model of dispositional attribution in interpersonal perception.” *Psychological Review*, 86: 61-79.

Richins, M. L. (1984). “Word of mouth communications as negative information.” In T. C. Kinnear (Ed.), “Advances in consumer research,” Vol. 11, Provo, UT: Association for Consumer Research.

Roodman, D. (2009). “Estimating Fully Observed Recursive Mixed-Process Models with cmp.” Center for Global Development, Working Paper 168.

Rosenbaum, M. E. (1986). “The repulsion hypothesis: On the nondevelopment of relationships.” *Journal of Personality and Social Psychology*, 51: 1156–1166.

Rosenberg, M. (1979). “Conceiving the self.” New York: Basic Books.

Rozin, P. and Royzman, E. B. (2001). “Negativity bias, negativity dominance, and contagion.” *Personality and Social Psychology Review,* 5: 296-320.

Ryan, R. M., Koestner, R., and Deci, L. (1991). “Varied forms of persistence: When free-choice behavior is not intrinsically motivated.” *Motivation and Emotion*, 15: 185-205.

Semmens, J. (1992). “Do seat belt laws work?” *The Freeman*, 42(7), available at http://www.thefreemanonline.org/columns/do-seat-belt-laws-work/.

Silver, R. L., Boon, C., and Stones, M. H. (1983). “Searching for meaning in misfortune: Making sense of incest.” *Journal of Social Issues*, 39, 81-102.

Simmons, J. P., Nelson, D. L., and Simonsohn, U. (2011). “False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant.” *Psychological Science*, 22: 1359-66.

Simon, H. A. (1982). “Models of bounded rationality.” MIT Press, Cambridge.

Singer, E., Hippler, H.J., and Schwarz, N. (1992). “Confidentiality assurances in surveys: Reassurance or threat?” *International Journal Of Public Opinion Research*, 4:256-68.

Sirgy, M. J. (1982). “Self-concept in consumer behavior: A critical review.” *Journal of Consumer Research,* 9(3): 287-300.

Skowronski, J. J. and Carlston, D.E. (1987). “Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases.” *Journal of Personality and Social Psychology*, 52: 689-699.

Skowronski, J. J. and Carlston, D. E. (1989). “Negativity and extremity biases in impression formation: A review of explanations.” *Psychological Bulletin*, 105: 131-42.

Slovic, P. (1987). “Perception of risk.” *Science*, 236(4799):280-285.

Slovic, P. (2010). “The feeling of risk.” Routledge, London, U.K.

Smith, H. J., Milberg, S. J., and Burke, S. J. (1996). “Information privacy: Measuring individuals’ concerns about organizational Practices.” *MIS Quarterly*, 20(2):167-196.

Solove, D. J. (2006). “A taxonomy of privacy.” *University of Pennsylvania Law Review*, 154(3):477-560.

Solove, D. J. (2007). “The future of reputation – gossip, rumor, and privacy on the internet.” Yale University Press, New Haven & London.

Stahl, D. and Wilson, P. (1994). “Experimental evidence on players’ models of other players.” *Journal of Economic Behavior and Organization*, 25: 309-327.

Steele, C. M. (1988). “The psychology of self-affirmation: Sustaining the integrity of the self.” In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 21, pp. 261-302). New York: Academic Press.

Stigler, G. J. (1961). “The economics of information.” *Journal of Political Economy*, 69(3): 213-225.

Styron, T., and Janoff-Bulman, R. (1997). "Childhood attachment and abuse: Long-term effects on adult attachment, depression and conflict resolution." *Child Abuse and Neglect*, 21: 1015-1023.

Strater, K. and Lipford, H. R. (2008). "Strategies and struggles with privacy in an online social networking community." *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction*. Volume 1: 111-119.

Stutzman, F. (2006). "An evaluation of identity-sharing behavior in social network communities." *Journal of the International Digital Media and Arts Association*, 3(1), 10-18.

Suri, S., Goldstein, D. G., and Mason, W. A. (2011). "Honesty in an online labor market." *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence.*

Sweeney, L. (1997). "Weaving technology and policy together to maintain confidentiality." *Journal of Law, Medicine & Ethics*, 25(2&3): 98-110.

Tavani, H. T. and Moor, J. H. (2001). "Privacy protection, control of information, and privacy-enhancing technologies." *Computers and Society*, 31(1): 6-11.

Taylor, S. E. and Brown, J. D. (1988). "Illusion and well-being: A social psychological perspective on mental health." *Psychological Bulletin*, 103(2): 193-210.

Taylor, S. E. and Mettee, D. R. (1971). When similarity breeds contempt. *Journal of Personality and Social Psychology*, 20(1): 75-81.

Tesser, A. (2000). “On the confluence of self-esteem maintenance mechanisms.” *Personality and Social Psychology Review*, 4: 290-299.

Tetlock, P. E. (1985). “Accountability: A social check on the fundamental attribution error.” *Social Psychology Quarterly*, 48: 227-236.

Trope, Y. (1979). “Uncertainty-reducing properties of achievement tasks.” *Journal of Personality and Social Psychology*, 37(9): 1505-1518.

Viard, V. B. and Economides, N. (2011). “The effect of content on global Internet adoption and the global 'digital divide.'” NET Institute Working Paper No. 10-24. NYU Law and Economics Research Paper No. 10-55.

Walker, K. (2000). “’It’s difficult to hide it:’ The presentation of self on Internet home pages.” *Qualitative Sociology*, 23(1): 99–120

Wang, Y., Komanduri, S., Leon, P.G., Norcie, G., Acquisti, A., and Cranor, L.F. (2011). “I regretted the minute I pressed share: A qualitative study of regrets on Facebook.” *7th Symposium on Usable Privacy and Security*, Pittsburgh, PA, USA.

Wardlaw, M. J. (2000). “Three lessons for a better cycling future.” *British Medical Journal*, 321(7276): 1582-1585.

Watson, D., Clark, L. A. and Tellegen, A. (1988). “Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales.” *Journal of Personality and Social Psychology*, 54(6): 1063-1070.

Weinstein, N. D. (1984). "Why it won't happen to me: Perceptions of risk factors and susceptibility." *Health Psychology*, 14: 431-457.

Weisband, S. and Kiesler, S. (1996). "Self-disclosure on computer forms: Meta-analysis and implications." *Proceedings of the SIGCHI Conference on Human factors in computing systems*, Vancouver, Canada.

Weiss, E. L., Longhurst, J. G., and Mazure, C. M. (1999). "Childhood sexual abuse as a risk factor for depression in women: Psychosocial and neurobiological correlates." *American Journal of Psychiatry*, 156: 816-828.

Westin, A. R. (1967). "Privacy and freedom." New York Atheneum.

Westin, A. R. and Harris Louis & Associates (1991). Harris-Equifax Consumer Privacy Survey. Technical report, Conducted for Equifax Inc.

Xu, H. (2007). "The effects of self-construal and perceived control on privacy concerns." *Proceedings of 28th Annual International Conference on Information Systems*, Montréal, Canada.

Zajonc, R. B. (1980). "Feeling and thinking: Preferences need no inferences." *American Psychologist,* 35: 151-75.

Zhao, S., Grasmuck, S., and Martin, J. (2008). "Identity construction on Facebook: Digital empowerment in anchored relationships." *Computers in Human Behavior*, 24(5): 1816-1836.

Zuckerman, M. and O'Loughlin, R.E. (2006). "Self-enhancement by social comparison: A prospective analysis," *Personality and Social Psychology Bulletin*, 32(6): 751-60.