APPLIED RESEARCH LAB
FOR INTELLIGENCE
AND SECURITY

Literature Review of the Relationship between Representations of the Self on Social Media and Representations of the Self Offline
Final Report

September 30, 2021

Long Doan[1]*, Paige Miller[1]
[1] Department of Sociology, University of Maryland, College Park
*corresponding author, longdoan@umd.edu

EXECUTIVE SUMMARY

Social media usage has been and continues to be on the rise (Auxier and Anderson 2021). Corresponding to this growth is increasing interest in better understanding how to use user-generated content like social media posts to infer information about the users that posted this content. Organizational scholars seek to understand how social media posts can foretell potential insider threats (Legg et al. 2013). However, little is known about the relationship between online content and offline selves. There are compelling reasons to expect that online and offline selves are different (Kozyreva et al. 2020). Online spaces are inherently different from and more constrained in many ways than offline spaces. Accordingly, unexplored discrepancy between online and offline selves may lead some factors to have inflated correlations to risk while other factors may have depressed correlations. Reviewing the work on the relationship between online and offline selves allows programs to better evaluate social media information and correct for distortions in the relationship between online selves and risk. To do this, we provide:

1. A systematic overview of the existing research across security, psychology, business, organizations, sociology and computer science disciplines,
2. Translation and synthesis of these disparate strands of research, and
3. Identification of feasible research questions that remain unanswered given this cumulative knowledge.

Three main findings emerge from our systematic review of the literature on online versus offline selves:

1. Online content can be used to determine valuable information about people's personalities,
2. Online self-presentations are generally accurate, and instances of deception are limited in scope and are often reactions to perceived barriers to social interactions rather than a desire to deceive, and
3. Online behaviors influence offline behaviors and vice versa.

However, there is emerging evidence as well as well-documented pitfalls with that comes with any operationalization of these findings. Specifically, many studies in this literature rely on self-selected and convenience samples to draw inferences and either treat demographic differences as unimportant or nonexistent. However, there is evidence that online users are not a monolithic group and there are indeed important demographic differences among users. For example, the research shows that:

- Gender, race, culture, age, and cohort all create "digital divides" that shape how and how much information people post online
- There are large personality differences in the meaning and purpose of social media posts

As such, using social media information, no matter how accurate, for personnel vetting and risk assessment should be balanced with the very real and easy to imagine ethical and practical dilemmas:

- Personnel need to provide consent to their social media information being used, potentially against them. Although this consent is relatively easy to obtain from the federal workforce,

consent is harder, if not impossible, to obtain from those who are connected, knowingly or unknowingly, to the focal person. This presents at least two problems:
    - Studies demonstrating the effectiveness of social media information in predicting personality are based on complete data. Censored data provided to the government due to the lack of third-party consent may be limited in predictability at best and even potentially misleading.
    - even if the data is adequately censored by vendors collecting social media information before passing them on to the government, it is difficult to monitor what these companies can and will do with the raw, uncensored data.
- There is a large potential for the unequal policing of certain subgroups compared to others. Existing data suggests that there will be strong differences in the likelihood of some groups being flagged by a risk assessment tool. Given that two people are of equal risk of being a malicious insider, the one who posts more frequently on social media will be more likely to be detected than the one who posts less frequently.
- Studies demonstrating that social media information can be used to predict personality, including "dark" personalities shown to be predictive of insider risk, are based on population aggregates. It is a stretch, then, to apply these findings to assessment and detection tools designed to predict individual-level risk.
- Machine learning and artificial intelligence, although predictive, cannot provide clear and convincing rationales for flagging someone as high risk. Black-box theorizing based on outputs leads to post-hoc explanations that are neither satisfying from an academic perspective nor concrete enough from a legal perspective. An algorithm can flag someone as being high risk, but if it cannot tell us *why* that person is high risk, it is of limited value.

Given the patterns identified in our literature review, we propose several key directions for future research:

1. Gaining a deeper understanding of subgroup differences and how various groups differentially use social media is needed to better contextualize and understand research findings.
2. The rise of social networks like Instagram and others focused on pictures and videos presents challenges that need to be addressed. Whether and how people represent themselves on these newer platforms compared to both more text-based platforms and offline interactions needs to be better understood.
3. The rise of "alternative" social media platforms that are not technically different but caters to specific subgroups as opposed to the general population, like Parlor, also complicates the relationship between online and offline selves.

Contents

INTRODUCTION

As social media usage continues to rise (Auxier and Anderson 2021; Hampton et al. 2011), there is increasing interest across a variety of disciplines in better understanding what information user-generated content like social media posts can tell us about the users that generated them. Scholars in fields like business and marketing are interested in using social media data to better understand purchasing decisions (Kim et al. 2012), while psychologists are interested in inferring stable personality traits from users' posts (Kosinski et al. 2014). Organizational scholars seek to understand how social media posts can foretell potential insider threats (Legg et al. 2013). Yet relatively little is known about the relationship between what individuals choose to post in online spaces and how that information is similar to or different from what they otherwise would have revealed about themselves offline. Existing insider threat programs use social media information as one source of data for generating risk scores for personnel (Pressman 2015). Yet, there are reasons to expect that online and offline selves are different (Kozyreva et al. 2020; Suler 2005). This discrepancy may lead some factors to have inflated correlations to risk while other factors may have depressed correlations. Reviewing the work on the relationship between online and offline selves allows programs to better evaluate social media information and correct for distortions in the relationship between online selves and risk.

In this report, we provide: (1) a systematic overview of the existing research across security, psychology, business, organizations, sociology and computer science disciplines, (2) translate and synthesize these disparate strands of research, and (3) identify feasible research questions that remain unanswered given this cumulative knowledge. In doing so, this review provides a roadmap toward generating and testing hypotheses aimed at understanding how social media and other online information can be used as a window into individuals' offline personas, behaviors, and their "true" selves. We begin by briefly reviewing the state of the published research on insider threat and assessment tools for malicious and non-malicious actors. Then, we detail the approach we used in this report to systematically review the existing literature studying online compared to offline selves. Next, we present major themes identified in our review of the literature, propose ways in which these findings can be put into practice, and pose unanswered questions based on this review that should be addressed.

INSIDER THREAT VS. INSIDER RISK

Work on insider threat has increasingly recognized the complexity in determining levels of threat and how to best detect this threat (Legg et al. 2013). Reflecting this recognition is a move from determining and categorizing threats to understanding and assessing risk (Pressman 2015). Much of the work in this domain has focused on conceptualizing and defining different types of insiders, their potential for producing damage, and how they can be detected, and the damage mitigated (Ho and Lee 2012; Ho and Warkentin 2017; Homoliak et al. 2018; Theoharidou et al. 2005). Although early work in this domain has often focused on case studies to examine malicious insiders' motivations and attack methodologies (Randazzo et al. 2004; Lynch 2006; Claycomb et al. 2012; Ross et al. 2009; Carter and Carter 2011; see BaMaung et al. 2018 for a recent review), more recently empirical research has moved to developing and testing different detection methods and

frameworks (Azaria et al. 2014; Axelrad et al. 2013; Ho et al. 2015; Kandais et al. 2013). Across these studies, scholars often delineate between malicious and non-malicious actors, and develop methods to assess the levels of threat associated with each type of actor (Legg et al. 2013).[1]

Risk Assessment Tools for Malicious Actors

Drawing on psychological insights that people's preferences and behaviors can partially be explained by personality traits, some have pushed for technological monitoring of insiders to flag potentially problematic personality traits for further review (Schultz 2002; Nurse et al. 2014; Brdiczka et al. 2012). The underlying premise of these approaches is that knowledge of underlying personality traits can reveal behavioral tendencies (Kosinski et al. 2014). A common approach is to detect insiders' level of Openness, Consciousness, Extroversion, Agreeableness, and Neuroticism (the OCEAN model) in addition to changes in behavior and motivations to detect potential insider threats (Legg et al. 2013). As a risk assessment tool, proofs of concept have shown that website browsing behaviors can be correlated with OCEAN to create risk profiles (Alahmadi et al. 2015). Along these lines, higher neuroticism and lower agreeableness and conscientiousness have been shown to relate to greater insider threat potential, while lower openness was linked to malicious potential (DuPuis and Khadeer 2016).

In contrast to creating profiles of potential insider threats, Pressman (2015) developed the Risk Assessment for Insider Threats (RAIT) tool that focuses on the situational and contextual factors that make an insider more likely to become a threat. The underlying premise of RAIT is the threat people pose as an insider is always changing because their circumstances change, so dynamic measures are needed, and assessment must be consistent. Using literature and case studies, Pressman (2015) identifies ten major risk indicators and twenty-five discrete indicators to detail the political, social, personal, economic, and protective factors someone might experience to increase or decrease their risk of becoming an insider threat.

Non-Malicious Actors

Although many studies implicitly or explicitly presume that insider threats are malicious in the harm they cause to organizations, non-malicious actors can also pose a threat to organizations. Work in this domain shows that although many of the personality traits are the same between malicious and non-malicious actors (DuPuis and Khadeer 2016), the tools needed to identify them are different. Because a common type of non-malicious actors are group dissenters (Packer and Chasteen 2010; Packer 2018), who deviate from or disagree with group norms, there may not be associated circumstantial changes to identify non-malicious insider threats (Colwill 2009). Instead, risk assessment tools for non-malicious actors tend to require more automated techniques (Gavai et al. 2015; Legg et al. 2015; Azaria et al. 2014).

[1] Malicious and Non-Malicious actors are trusted members of an organization, who harm their organizations with their access to sensitive information. Malicious actors typically cause harm by using the information for personal gain, while non-malicious actors do so unintentionally by misusing systems or not adhering to security protocols and may not even be aware of the breach (Dupuis and Khadeer 2016).

Monitoring Online Activities

Because many traits overlap between malicious and non-malicious insider threats (DuPuis and Khadeer 2016), and the damage caused by malicious and non-malicious actors can be equally consequential (Pfleeger 2008), much of current work in this literature has shifted to mass surveillance of online activities as a potential solution to assessing all insiders' level of risk to the organization (Gritzalis 2014). As the field moves from individual case studies of insider threats to more holistic and automated mass gathering of online data (Greitzer and Frincke 2010), it is important to systematically understand how and under which conditions online information is reflective of people as a whole. There are many reasons to expect that online and offline behaviors and selves may differ (Kozyreva et al. 2020; Suler 2005). Kozyreva and colleagues outline several systematic differences between online and offline environments that may alter individuals' perceptions, preferences, and behaviors. Group size and friendship networks tend to be larger online than offline (Dunbar 2016); information is more readily available online (Schwartz 2016); online environments change more rapidly than offline environments (Roberts 2018); online environments are more personalized than offline environments (Burrell 2016); and online environments are mediated by user interfaces that restrict users' choices (Berners-Lee et al. 1992; DeAndrea and Walther 2011).

Given these structural differences, it is likely that people are not exactly the same online, or even received the same, as they would be offline (Okdie et al. 2011; Rouse and Haas 2003; Suler 2005) and this may change over time (Yang and Brown 2016). Yet, there are relatively little systematic examinations comparing people's online selves to their offline counterparts. Such an examination is necessary to contextualize and evaluate risk assessment tools using such information as its input. Without knowing when, how, and why people's online selves are different from their offline selves, such tools can be of limited utility at best and counterproductive at worst. Below, we describe the approach we take to our systematic literature review and key insights from it.

THEORIES OF THE SELF OFFLINE AND ONLINE

Moving from offline environments to online environments leads to what scholars call a "context collapse"–the compression of once stable and separate networks and environments onto a singular platform by the internet (Davis and Jurgenson 2014). As such, findings that appear to conflict with one another at face value, may fit into a larger pattern of human behavior and motivations if contexts were considered. Although not part of this systematic review, we provide a brief overview of the main theories of the self to help make sense of the broader findings we present later. Because online platforms are inherently social arenas, this review focuses on social conceptions of the self to understand online behavior.

Social psychological theories have long accepted the mutual influence of the person and society on selfhood. Mead (1934) made this distinction by separating the self into the "I" and the "Me," with the "Me" accounting for the social aspects of the self and the "I" representing impulses. The self is both the subject (I) and the object (me). Cooley (1902) goes further and proposes that people form a sense of self through a "looking glass," which is to say that people often reflect or think of

themselves, how they believe others think of them. Goffman's (1959) dramaturgical theory of the self, adding that not only do others participate in ourselves, but we also anticipate and actively perform to the idea of self we wish to maintain so that it might be validated and reaffirmed by others, calling these phenomenon "impression management" and "facework." This is not to say that people pretend to be themselves, but rather that work goes into being seen in the way one wants to. A part of "facework" is the front stage and backstage, which refers to the active and preparatory stages of interaction. Stryker expanded on this line of thought with identity theory (Stryker and Burke 2000). In addition to society and socialization shaping the person and their behavior, the context, associated roles, and relationships tied to specific identities can affect which identities become evoked in a given situation.

For each of these theories, the self is collective, reflexive, and sometimes strategic. Thus, emphasizing the importance of contexts and social networks in understanding the self and behavior. On its face, these theories suggest that the structural differences between online and offline environments would lead to different selves online compared to offline. The affect control theory of the self (Heise and MacKinnon 2010) suggests that people who hold identities that are not reflective of their self-image may seek to offset the meanings of these identities in other domains of their lives. For example, if one holds a job that conflicts with one's self-image, one may take on hobbies that counteract the meanings of that job. However, these theories of self also emphasize the stability and relative importance of some identities over others within the self (Stryker and Burke 2000), so perhaps people carry their identities across contexts when they can. Given this core understanding of the self and people's strategic management of their selves, it is unclear if online environments would lead to drastically different self-presentations or an extension of their offline selves. We seek to explore the empirical literature to find out.

LITERATURE REVIEW METHOD

Following best practices recommendations for systematic literature review (Munafò et al. 2017), our research team determined key parameters for the literature search and pre-registered our approach on the Open Science Framework (OSF) platform (available at https://osf.io/7s62n). Our research plan proceeded in two phases. In the first phase, we defined the basic inclusion and exclusion criteria for the literature review, identified *a priori* keywords to begin our search, determined the conditions under which new criteria could be included and how existing criteria could be modified, and basic data hygiene conventions. To end the first phase, we conducted an initial search as a pilot for the full literature review. In the second phase, we implemented the pre-registered plan. In all, we identified 142 potential sources for inclusion. Of these, 95 fit our inclusion criteria and 47 did not. Below, we describe our pre-registered research plan that resulted in these articles and chapters being included.

Pre-Registered Research Plan

Drawing on our respective expertise, we initially decided that our review will focus on the "human factors" of insider risk, such as the behavioral, psychological, and environmental factors that lead to insider events. Thus, our plan was to exclude research on specific algorithms, data scraping

techniques, or particular technologies used to study online and offline selves. To the extent that research uses these algorithms and techniques, we will include them if they meet our other inclusion criteria. Broadly, we include all peer-reviewed, non-classified research that examines the relationship between the content and nature of online presentation of self and how that online self relates to insider risk and/or offline aspects of the self that relates to insider risk. In other words, we include articles that speak to insider risk either directly (e.g., types of online content that pose a risk) or through offline personality and demographic factors (e.g., types of online content that relate to personality traits shown in past work to pose a risk).

Within our inclusion criteria, we prioritized certain research areas. Studies that fit a prioritized research area have their reference lists checked for potential new sources that may have been missed by our keyword search. Non-prioritized research that fits our inclusion criteria are included, but their references are not checked. Inclusion priority was determined by how well content overlapped with the goals of the project. Our main areas of interest were in online and offline identities, the workplace, and the individual in relation to a group. Studies on social media use, how individuals present themselves online, and how similar or dissimilar these behaviors may be from the way one behaves face-to-face are important for evaluating how well they can indicate in-person behavior. Similarly, the workplace is a significant content area in this research because it is the main context where events leading up to and of insider threat take place. Boundaries between individuals' work and personal life, whether they have a sense of belonging or identify with the working group, organizational culture, and the presence of workplace discontent and burnout may set the stage for acting against the working community. Trust, loyalty, motivation, and revenge were included for a more personal and emotional approach to insider threat. Lastly, research on peer surveillance and masculinities were chosen for their ties to interpersonal violence and bystanding.

We initially searched for the following keywords: insider threat, intragroup dissent, intragroup conflict, malicious insider, threat management, online offline selves, online self-presentation, online offline behavior, social media self, social media presentation. After reviewing the first 20 sources that resulted from these keywords, we did not add more search terms as they captured the variety of research we were interested in finding. We instead focused our attention on finding sources from the reference list of prioritized studies. Of our 142 potential sources, 53 came from non-keyword reference list searches. The exclusion rate between keyword and reference list sources are similar (33 compared to 34 percent, respectively). We continued searching for sources until we both agreed that we have reached saturation with our literature review and are not learning substantially new insights from additional new sources.

Although we stuck to our pre-registered research plan throughout the literature review, we did make one necessary clarification to our research plan as we encountered sources that were ambiguous in terms of fitting our inclusion criteria. We decided that for book sources, it was too difficult, especially for edited volumes, to have a binary choice as to whether the entire book fit the inclusion criteria. Therefore, we decided that if a book had a singular thesis, we would determine if

the book as a whole fit the inclusion criteria, but that we would evaluate edited volumes at the chapter level.

SELF-PRESENTATION ONLINE

Existing research on online self-presentation is focused on three main overarching domains: 1) inferring personality traits from online content, 2) the selective presentation of self and the strategic use of deception online, and 3) how online and offline behaviors mutually influence one another. Across these three domains, the key takeaway is that, despite structural differences between online and offline environments, people tend to accurately represent themselves online (Bargh et al. 2002; Bortree 2005; McKenna 2007). Although there are clear examples and caveats to this conclusion, there is more evidence than not that online content is a useful and mostly accurate window into people's selves (Hogan 2010). We review this supporting evidence in more detail below.

Inferring Personality from Online Content

Perhaps most directly linked to the existing insider threat literature, one of the major domains of research in online selves is focused on using online content to infer psychological personality traits (Back et al. 2010; Balani and Choudhury 2015; Correa et al. 2010; Hocevar et al. 2014). Early work in this domain was focused on easily accessible information within organizations like email communications and blogs (Yarkoni 2010). In a content analysis of over 150 emails, Brown and colleagues (2013) found that email writing styles are associated with personality traits. People high in agreeableness, but low in neuroticism via email pose a lower risk of becoming a malicious insider. The reverse is also true, however, people who are high in both agreeableness and neuroticism pose more of a risk to their organization. People were also found to be at high risk if they tended to communicate negative emotions more frequently. Although email can be used to infer personality, some traits are easier to infer from written text than other traits (Gill et al. 2006). Among a sample of observers unaffiliated with email authors, there is higher agreement for extraversion, less so for psychoticism and neuroticism. Extraversion is also most accurate (observer rating matched self-report from email author). Neuroticism is negatively correlated with self-reports, suggesting that observers are bad at guessing that trait. There is also individual level variation in who is better at guessing which trait. Extending this line of inquiry, researchers are deploying machine learning methods to process more emails (Alahmadi et al. 2015) and combining emails with other organizational information like file access patterns to infer personality traits and the level of risk associated with such traits (Gavai et al. 2015).

Beyond the use of emails and other organizational data to infer personality traits, researchers are using big data techniques to collect and analyze website usage and social media data to infer personality traits. In one of the first big data analyses of online presentation of self, Kosinski and colleagues (2014) showed that information that Facebook users choose to present about themselves on their profile pages as well as the types of websites they frequented both independently and jointly predict their big five personality traits. More research in the area based on questionnaires, found that people higher in neuroticism and lower in self-esteem had higher

rates of online activity overall (Mehdizadeh 2010). Unlike many smaller scale studies of this ilk from diverse samples like Canadian MPs (Koop and Marland 2012) to website owners (Marcus and Schuler 2004; Marcus et. al 2006), the big data analyses of website and social media data rely on machine learning to identify patterns of correlations between aspects of online selves and offline personality traits rather than existing theory (Papacharissi 2002). Indeed, Kosinski's study acknowledges this "black box" theorizing and points to patterns of correlations that may not make theoretical or intuitive sense but are nonetheless predictive.

In addition to mining Facebook profile data (Golbeck et al. 2011; Kosinski et al. 2014; Michikyan, Subrahmanyam, and Dennis 2014; Ong et al. 2011; Back et al. 2010), scholars have been able to show that useful personality information can be gleaned from shorter content like Twitter tweets (Golbeck et al. 2011) within 11-18% of users' actual Big Five Personality scores. However, more recent work warns of an overoptimistic assessment of studies aiming to and successfully showing the social media usage patterns can be predictive of people's personality traits. Sumner and colleagues (2012) show that although machine learning algorithms do a much better job than chance at predicting personality *aggregating across thousands of profiles*, even the most successful algorithm only has about a 60 percent accuracy rate for any individual profile. In other words, the promise of big data in predicting personality traits is in determining how various populations' social media profiles and posts are associated with personality traits, not how individuals' social media profiles and posts are associated with their own personality traits. We may know that people who tend to use parentheses in their tweets are less open and extroverted than people who do not (Golbeck et al. 2011), but we cannot use that trend to say that I am less open and extroverted than my friend (because I use parentheses more often than she does).

Deception and Authenticity Online

A second stream of research asks the question: when and why do people present a deceptive picture of themselves online? Within this domain of research, scholars typically focus on two major topics: 1) identifying characteristics of people who view online spaces as a way to present an idealized version of themselves to improve their self-image and 2) how dating environments represent a unique nexus of online-offline self-presentations. Across these topics, it is clear that many people present an accurate picture of themselves online (DeVito et al. 2017; DeVito et al. 2018). Deception that exists in online self-presentations tend to be exaggerations and flourishes rather than outright lies. However, there is variation in who chooses to present a deceptive picture of themselves.

One of the primary factors shaping frequency and content of online interactions is whether someone is introverted or extroverted (Gosling et al. 2011; Krämer and Winter 2008). Gosling and colleagues analyze the frequency and content of Facebook posts of self-identified introverts and extroverts to find that extroverts are more likely to seek out virtual engagements and use the site more frequently than introverts. They use this finding to argue that people tend to extend their offline presence to online spaces rather than compensating for them (Tosun and Lajunen 2010). Although introversion and extroversion play out similarly online as they do offline, research also

shows that introverted people are more likely to view their online personas as a more accurate representation of their "true selves" than their offline personas (Amichai-Hamburger et al. 2010). Amichai-Hamburger and colleagues find that introverts take advantage of the relative anonymity online combined with more control over what they present about themselves and how to do it as primary drivers enabling introverts to present a more accurate picture of themselves online. This may be part of the reason that introverts are more likely to prefer online interactions while extroverts are more likely to prefer offline interactions (Goby 2006).

Studies also point to how people's social characteristics shape not only how they use social media, but *why* they use social media (Zywica and Danowski 2008; Hayes et al. 2015). Zywica and Danowski find people engage in various forms of social enhancement, but for different reasons. They find that individuals with lower self-esteem and high sociability were more likely to enhance their presentations online to appear popular or an idealized version of themselves. (Offline) popular people tend to want to maintain their offline status online via enhancement while (offline) unpopular people tend to want to gain online status via enhancement online. In addition, scholars find that people who view their "real selves'' closer to their online selves are more likely to value online interactions (Marriott and Buchannan 2014). This suggests that people's definitions of themselves affect the way in which they approach online versus offline interactions (Widyanto and Griffiths 2011). However, a growing body of research focuses on online interactions that have the potential to become offline interactions: online dating.

Online dating is different from other online representations of selves because people go in with the intention of meeting in person so there's more incentive to faithfully represent themselves (Gibbs et al. 2006). If so, this is an ideal space to look for online versus offline concordance. By interviewing online daters about how and why they present various versions of themselves online, Ellison and colleagues (2006) show that people strategically use deception to "get their foot in the door," but are generally focused on presenting themselves accurately. They find that people try to balance true representations of themselves with presenting an idealized version. They are very aware of how they might be perceived and "work around" technical barriers to make sure they are not filtered out. For example, they value giving an accurate overall impression of themselves, but they sometimes outright lie about characteristics like age because they don't want to be filtered out in searches. They also focus on subtle cues in others' profiles to form impressions of them.

In general, less attractive people are more likely and incentivized to present a deceptive picture of themselves (Toma et al. 2010), but the deceptions are still correlated with the truth because people need to balance the deception with the potential of future face to face interactions (Toma et al. 2008). There is some evidence that men are more likely to exaggerate their desirable features in dating profiles than women (Guadagno et al. 2012), but other studies do not find this gender difference (Toma et al. 2010). Much of literature on dating profiles show that deception in self presentations are often people's attempts to adapt to specific quirks of the dating platforms they use (Ward 2016). In Ward's analysis of Tinder, findings suggest that people carefully select pictures that make them seem desirable and disclose facts about themselves like layers of an onion. They

also use other people's profiles to figure out how to best present themselves, suggesting that a platform's social norms play an important role in shaping deception online net of user intentions.

Mutual Influence of Online and Offline Behaviors

Studies on human behavior and online environments typically focus on the person behind the screen. What can user generated content say about the way the person behaves face-to-face? Personality, for instance, can change how users navigate online spaces: what sites they use and how they interact on them (Goby 2006). This would suggest the scale tips in one direction: the offline existence or circumstances of the internet user imposes upon the online environment. There is ample evidence to support this thesis. Internet users by and large choose to be friends with their face-to-face friends online (Rui and Stefanone 2013) and choose to recreate their offline lives online through avatars or simulation games (Linares et al. 2011). Not only do users often recreate themselves online (Wynn and Katz 1997, but they rarely make significant alterations to their presentation of self online (Bullingham and Vasconcelos 2013). People who report stealing face-to-face even report stealing online at similar rates (Ogan, Ozakca, and Groshek 2008). A natural experiment conducted by Falavarjani and colleagues found that changes in offline behavior was able to strongly predict a following change in behavior online (2019). Face-to-face motivations for unfriending a friend on social media are likewise considered the most powerful and permanent motivators by users to end relationships (Sibona and Walczak 2011). It's not a stretch to see the ways that the interior, offline, circumstances of a person's life can impact behavior online.

However, there is just as much evidence to suggest the scales are more evenly balanced between online and offline influence, showing how online activity can impact face-to-face behavior. Research on online simulation games presents growing evidence of a phenomena called "the proteus effect," where the appearance of an online avatar has the ability to influence in-game behavior and even future face-to-face interactions (Yee, Bailenson, and Ducheneaut 2009). New mothers can use social media to mitigate isolation, improve self-esteem, and receive social support (Djaforova and Trofimenko 2017). Community-based discussion boards improve local civic engagement and neighborhood safety (Erete 2015). This is where the negative impacts of online environments start to be seen as well. The social comparison that social media affords can have very negative consequences on user body image and perceived attractiveness (Fox and Vendemia 2016). Women, and young women in particular, consistently report the detrimental effect of social media use on their well-being and self-esteem (Vogel et al. 2014; Vogel and Rose 2016). In fact, some studies have found that looking at one's own social media profile was more impactful to self-esteem than looking in a mirror. (Gonzales and Hancock 2011). These findings serve as a reminder of the dangers of self-objectification and the treatment of the self as text online.

All of this highlights the mutual influence of online and offline user interactions. As users continue to shape technology and social networks, they shape the life of the user offline. A study of adolescents' online activity showed that they were likely to use the internet to cope with loneliness in ways consistent with their offline coping mechanisms (Seepersad 2004). Young adults who viewed content showing risky or unsafe behaviors were more likely to take part in risky behaviors

offline (Branley and Covey 2017). As the portion of the population who lived before the internet declines and Americans spend more time online (Hampton et al. 2011), this relationship is expected to strengthen.

MOVING FORWARD

Three main findings emerge from our systematic review of the literature on online versus offline selves: 1) online content can be used to determine valuable information about people's personalities, 2) online self-presentations are generally accurate and instances of deception are limited in scope and are often reactions to perceived barriers to social interactions rather than a desire to deceive, and 3) online behaviors influence offline behaviors and vice versa. In this section, we translate these lessons into practical implications for monitoring insider threat and propose fruitful areas for future research. We first begin with emerging evidence that temper our findings. Specifically, many studies in this literature rely on self-selected and convenience samples to draw inferences and either treat demographic differences as unimportant or nonexistent. However, there is evidence that online users are not a monolithic group and there are indeed important demographic differences among users.

Demographic Differences Online

Research shows that users don't exist as a monolithic group and the differences between them are just as important as the differences outside, such as to non-users. Elijah Cassidy highlights this by showing how sexual minorities navigate online spaces to expand social networks and vet potential dates (2013). Cassidy's work demonstrates what they coin the "privacy divide," which is to say that not all groups of internet users have the same expectation or access to privacy and safety online. This sentiment is echoed in research on online dating between heterosexual couples. Women on dating sites were more likely to want to exchange more messages before going on a date with men, while men were more likely to introduce deception and exaggerations when they thought they would meet women for an in-person date (Gaudagano, Okdie, and Kruse 2012; Okdie et al. 2011).

The differences continue across arenas and groups. Women were more likely to be on social networking sites, more educated people were more likely to be on Facebook and LinkedIn, and the largest demographic increase of people online were in people who are over 35 (Hampton et al. 2011). However, the largest differences reported in research fall along gender (Correa et al. 2010; Fox and Rooney 2015). Women are found to use social networks more for comparison, while men tend to use the networks to make friends (Haferkamp et al. 2012). In addition, women put more effort into posts due to social expectations about image and desirability. They were also more likely to feel negatively after the social comparison (Fox and Vandemia 2016). Demographic pockets on certain sites and differential treatment sets the stage to an altogether different experience of the internet.

It should be no surprise, then, that generations raised with the internet would have more novel interactions on it. Studies of adolescent online behaviors overwhelmingly show that younger groups used social media to supplement their existing offline relationships (Subrahmanyam et al

2008), deal with loneliness online in ways that were consistent with their face-to-face coping mechanisms (Seepersad 2004), and match their online selves to their offline selves (Calvert et al. 2003). Overall, students who spend more time online reported a more authentic online self (Michikyan, Subrahmanyam, and Dennis 2014). This is to say that for groups who spend a greater portion of their lives online, the consequences of online spaces are more significant and real because it is a larger part of their reality and day-to-day experience.

While demographic differences in online behavior are important to understanding how experiences of the internet differ, they should be taken with caution. Many internet studies focus solely on gender and age for comparison and use university students as their sample without larger or more meaningful reference groups to make sense of the data. They are also primarily based on samples from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies (Henrich et al. 2010). Findings, and understandings of online life, then, are skewed to relatively advantaged experiences, who are operating on the internet under certain assumptions of access and safety. The findings detailed in this section, for instance, may not hold true for more disadvantaged and older groups. Though studies are beginning to acknowledge culture and race as important determinants of online experience (Wang et al. 2012; Rui and Stefanone 2013), there still exists an unfortunate lack of data to demonstrate these differences in greater and finer detail.

Operationalizing the Research

Although the research on online selves show that social media representations of selves include valuable information for personnel vetting and determining insider risk, creating concrete policies from this research is less straightforward. In this section, we outline several key considerations based on our assessment of the literature and potentially fruitful directions for future research. These operationalization considerations broadly concern issues of ethics and consent, the potential for unequal policing of social media content, and the cost to benefit of collecting online information.

Perhaps the biggest barrier to using social media information for vetting and risk assessment concerns issues of ethics and privacy. Research shows that public discourse on invasions of privacy tend to be stronger and more negative toward government compared to corporate actors even when the act itself is similar in nature (Connor and Doan 2021), which suggests that any backlash from such a program needs to be carefully weighed against the value of any potential information learned. Regardless of the public reaction, personnel need to be aware and provide consent to their social media information being used, potentially against them. Although this consent is relatively easy to obtain from the federal workforce, consent is harder, if not impossible, to obtain from those who are connected, knowingly or unknowingly, to the focal person. At least two major issues arise from the lack of consent from associates of the focal person.

First, studies demonstrating the effectiveness of social media information in predicting personality are based on complete data. Scholars scrape all publicly available information and build their machine learning algorithm based on these data. Censored data provided to the government due to the lack of third-party consent may be limited in predictability at best and even potentially

misleading. We simply do not know, based on extant research, if censored data are predictive in the same way as complete data. Second, even if the data is adequately censored by vendors collecting social media information before passing them on to the government, it is difficult to monitor what these companies can and will do with the raw, uncensored data.

Related to the issue of consent and ethics is the likely potential for the unequal policing of certain subgroups compared to others. We know that extroverted people, for example, are more likely to post content than introverted people (Gosling et al. 2011). There are also generational differences in frequency and mode of social media usage (Subrahmanyam et al. 2008). And scholars are only beginning to look at cultural differences online (Rui and Stefanone 2013). Given that the limited data that exists suggests key demographic and personality differences in how often someone uses social media (Correa et al. 2010), these data suggest that there will be strong differences in the likelihood of some groups being flagged by a risk assessment tool. In other words, given that two people are of equal risk of being a malicious insider, the one who posts more frequently on social media will be more likely to be detected than the one who posts less frequently. Any assessment and detection algorithm needs to account for these denominator differences in order to be equitable in its application.

Finally, it is important to weigh the potential value of social media information with the cost of collecting and analyzing them. Doing so requires that we acknowledge the difference between higher propensity of risk and actually being a risk. Two issues are clear in the literature that makes the usage of social media information tricky. First, studies demonstrating that social media information can be used to predict personality, including "dark" personalities shown to be predictive of insider risk, are based on population aggregates. These studies show that on average, people who tend to post in certain ways tend to have related personality traits (Kosinski et al. 2014), not that any given individual will have this tendency (Sumner et al. 2012). It is a stretch, then, to apply these findings to assessment and detection tools designed to predict individual-level risk. Second, machine learning and artificial intelligence, although predictive, cannot provide clear and convincing rationales for flagging someone as high risk. Black-box theorizing based on outputs leads to post-hoc explanations that are neither satisfying from an academic perspective nor concrete enough from a legal perspective. An algorithm can flag someone as being high risk, but if it cannot tell us *why* that person is high risk, it is of limited value.

With this context in mind, the cost-benefit analysis of collecting massive amounts of potentially invasive data on large swathes of the population seems less enticing than at first glance. Although we know that the information gathered is likely to be indicative of offline selves, relatively deception-free, and generally predictive, the ethical and practical costs of collecting and analyzing these data may outweigh their benefits, for now. If the data can be better understood and the risk of inequitable surveillance can be mitigated, the cost-benefit calculation will be more enticing.

Unanswered Questions

Given that the value of social media information to personnel vetting and risk assessment is dependent on the completeness and satisfactory explanation of why someone is flagged, we focus on fruitful directions for future research in this domain. Taking the literature as a whole, the most obvious direction for future research relates to the emerging evidence of demographic and personality differences in social media usage rates and patterns. Many studies, although large in size, rely on convenience samples and cannot be extrapolated to a broader population. It is unlikely that a representative sample will provide consent for their social media information to be mined for risk assessment. Therefore, a deeper understanding of subgroup differences and how various groups differentially use social media is needed to better contextualize and understand research findings.

Among the key demographic and personality differences are comparisons between introverts and extroverts (Amichai-Hamburger et al. 2010), gender (Haferkamp et al. 2012), and cohort differences (Subrahmanyam et al. 2008). These factors have already been shown to be linked to differential usage patterns and are worthy of further examination. Cohort differences in social media usage is among the strongest and most consistent differentiators in use. Alternatively, we know relatively little about cultural differences in social media usage (Rui and Stefanone 2013), and there is more potential for undiscovered differences by looking at factors that have not been explored in the literature. Studies in our review are primarily focused on the US and Americans. The studies we found that started looking at cultural differences are focused on US-Asian comparisons (Wang et al. 2012; Rui and Stefanone 2013). A focus on other cultural differences is needed as is an understanding of racial and ethnic differences in social media use within cultures.

As scholars begin to explore these unanswered questions, studies should continue to balance detection and ethics. Exploring demographic differences in social media usage patterns and potentially revealing that certain subgroups are more or less likely to post content that may be flagged by a risk detection algorithm raises the uncomfortable but inevitable question of what the line is between risk detection and profiling based on immutable characteristics. This is another reason by "black box" theorizing and focusing on algorithms that find predictive patterns without providing clear and convincing reasons why someone is flagged as being high risk is problematic.

Finally, almost all studies in our review are based on text analysis of social media information. The rise of social networks like Instagram and others focused on pictures and videos presents challenges that need to be addressed. Whether and how people represent themselves on these newer platforms compared to both more text-based platforms and offline interactions needs to be better understood. Relatedly, the rise of "alternative" social media platforms that are not technically different but caters to specific subgroups as opposed to the general population, like Parlor, also complicates the relationship between online and offline selves.

As shown in our review, the relationship between online and offline selves is strongly and positively correlated. However, that relationship is heterogeneous, and that heterogeneity should

be better understood. As newer and more targeted platforms grow in popularity, the research should also be updated to reflect changes in the rise of these platforms. Using this information for risk assessment requires careful consideration of both the risk and benefits, including very foreseeable ethical problems that need to be transparently addressed.

ACKNOWLEDGEMENTS

This report was prepared for Office of the Undersecretary of Defense for Intelligence and Security (OUSD(I&S)), United States Department of Defense (DoD) under the following agreement: HQ003420F0655, *University of Maryland,* "Insider Threat and Personnel Vetting."

DISCLAIMERS

Any views, opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of an official United States government position, policy, or decision. Additionally, neither the United States government nor any of its employees make any warranty, expressed or implied, nor assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, product, or process included in this publication.

Certain commercial entities, equipment, or materials may be identified in this document to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the Applied Research Laboratory for Intelligence and Security (ARLIS), the University of Maryland, or the United States government, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

ABOUT ARLIS

Applied Research Laboratory for Intelligence and Security (ARLIS) is a UARC based at the University of Maryland College Park and established in 2018 under the auspices of the OUSD(I&S). ARLIS is intended as a long-term strategic asset for research and development in artificial intelligence, information engineering, acquisition security, and social systems. One of only 14 designated United States Department of Defense (DoD) UARCs in the nation, ARLIS conducts both classified and unclassified research spanning from basic to applied system development and works to serve the U.S. Government as an independent and objective trusted agent.

Technical Points of Contact:

PI: Adam Russell, D.Phil.
Chief Scientist, ARLIS
301.226.8834; arussell@arlis.umd.edu

Task Lead: Long Doan, Ph.D.
Assistant Professor, Department of Sociology
301.405.7586; longdoan@umd.edu

Co-PI: Kelly Jones, Ph.D.
Assistant Research Scientist, ARLIS
301.226.8850; kjones@arlis.umd.edu

Administrative Points of Contact:

Ms. Monique Anderson
Contract Officer, Office of Research Administration
Assistant Director, ARLIS
301.405.6272; manders1@umd.edu

REFERENCES

Alahmadi, Bushra A., Philip A. Legg, and Jason RC Nurse. 2015. "Using internet activity profiling for insider-threat detection." *Proceedings of the 17th International Conference on Enterprise Information Systems, Volume 2*.

Amichai-Hamburger, Y., & Vinitzky, G. 2010. "Social network use and personality." *Computers in Human Behavior* 26:1289–1295.

Amichai-Hamburger, Y., Wainapel, G., & Fox, S. (2002). "On the internet no one knows I'm an introvert": Extraversion, Neuroticism, and internet interaction. Cyberpsychology and Behavior, 5(2), 125-128.

Auxier, Brooke, and Monica Anderson. 2021. "Social Media Use in 2021: A majority of Americans say they use YouTube and Facebook, while use of Instagram, Snapchat and TikTok is especially common among adults under 30." *Pew Research Center*. Available at https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/

Axelrad, Elise T., et al. 2013. "A Bayesian network model for predicting insider threats." *2013 IEEE Security and Privacy Workshops.* IEEE.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. 2010. "Facebook profiles reflect actual personality, not self-idealization." *Psychological Science* 21(3): 372-374.

Balani, Sairam, and Munmun De Choudhury. 2015. "Detecting and characterizing mental health related self-disclosure in social media." *Proceedings of the 33rd Annual ACM Conference.*

BaMaung, David, et al. 2018. "The enemy within? The connection between insider threat and terrorism." *Studies in Conflict & Terrorism* 41: 133-150.

Bargh, J. A., McKenna, K. Y. A., & Fitzsimons, G. M. 2002. "Can you see the real me? Activation and expression of the true self on the internet." *Journal of Social Issues* 58: 33-48.

Berners-Lee, T., Cailliau, R., Groff, J.-F., & Pollermann, B. (1992). World-wide web: The information universe. *Internet Research*, *20*, 461–471.

Bortree, Denise Sevick. 2005. "Presentation of self on the web: An ethnographic study of teenage girls' weblogs." *Education, Communication & Information* 5: 25-39.

Branley, Dawn Beverley, and Judith Covey. 2017. "Is exposure to online content depicting risky behavior related to viewers' own risky behavior offline?." *Computers in Human Behavior* 75: 283-287.

Brdiczka, Oliver, et al. 2012. "Proactive insider threat detection through graph learning and psychological context." *2012 IEEE Symposium on Security and Privacy Workshops*.

Brown, Christopher R., Alison Watkins, and Frank L. Greitzer. 2013. "Predicting insider threat risks through linguistic analysis of electronic communication." *2013 46th Hawaii International Conference on System Sciences*.

Bullingham, Liam, and Ana C. Vasconcelos. 2013. "'The presentation of self in the online world': Goffman and the study of online identities." *Journal of Information Science* 39: 101-112.

Burrell, J. 2016. "How the machine 'thinks': Understanding opacity in machine learning algorithms." *Big Data & Society* 3. https://doi.org/10.1177/2053951715622512

Calvert, Sandra L., et al. 2003. "Gender differences in preadolescent children's online interactions: Symbolic modes of self-presentation and self-expression." *Journal of Applied Developmental Psychology* 24: 627-644.

Carter, J. G. & Carter, D. L. 2011. "Law enforcement intelligence: implications for self- radicalized terrorism." *Police Practice and Research* 13: 138-154.

Cassidy, Elija M. 2013. *Gay men, social media and self-presentation: Managing identities in Gaydar, Facebook and beyond*. Dissertation. Queensland University of Technology.

Chinchani, Ramkumar, et al. 2005. "Towards a theory of insider threat assessment." *2005 International Conference on Dependable Systems and Networks*.

Claycomb, W. R. et al. 2012. "Chronological examination of insider threat sabotage: preliminary observations." *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 3: 4-20.

Colwill, Carl. 2009. "Human factors in information security: The insider threat–Who can you trust these days?" *Information Security Technical Report* 14: 186-196.

Connor, Brian T., and Long Doan. 2021. "Government and corporate surveillance: moral discourse on privacy in the civil sphere." *Information, Communication & Society* 24: 52-68.

Cooley, Charles Horton. 1902. *Human Nature and the Social Order*. Schribner's Sons: New York.

Correa, T., Hinsley, A. W., & Gil de Zuniga, H. 2010. "Who interacts on the Web? The intersection of users' personality and social media use." *Computers in Human Behavior* 26: 247-253.

Davis, Jenny L., and Nathan Jurgenson. 2014. "Context collapse: Theorizing context collusions and collisions." *Information, communication & society* 17: 476-485.

DeAndrea, D. C., & Walther, J. B. 2011. "Attributions for inconsistencies between online and offline self-presentations." *Communication Research* 38: 805–825.

DeVito, Michael A., et al. 2018. "How people form folk theories of social media feeds and what it means for how we study self-presentation." *Proceedings of the 2018 CHI conference on human factors in computing systems*.

DeVito, Michael A., Jeremy Birnholtz, and Jeffery T. Hancock. 2017. "Platforms, people, and perception: Using affordances to understand self-presentation on social media." *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*.

Djafarova, Elmira, and Oxana Trofimenko. 2017. "Exploring the relationships between self-presentation and self-esteem of mothers in social media in Russia." *Computers in Human Behavior* 73: 20-27.

Donath, J. 1999 "Identity and Deception in a Virtual Community," in M. Smith and P. Kollock (eds) *Communities in Cyberspace*, pp. 29–59. London: Routledge.

Dunbar, R. I. M. 2016. "Do online social media cut through the constraints that limit the size of offline social networks?" *Royal Society Open Science* 3: 150292.

Dupuis, Marc, and Samreen Khadeer. 2016. "Curiosity killed the organization: A psychological comparison between malicious and non-malicious insiders and the insider threat." *Proceedings of the 5th Annual Conference on Research in Information Technology*.

Ehrenberg, A., Juckes, S., White, K. M., & Walsh, S. P. 2008. "Personality and self-esteem as predictors of young people's technology use." *Cyberpsychology and Behavior* 11: 739–741.

Ellison, Nicole, Rebecca Heino, and Jennifer Gibbs. 2006. "Managing impressions online: Self-presentation processes in the online dating environment." *Journal of Computer-Mediated Communication* 11: 415-441.

Erete, Sheena L. 2015. "Engaging around neighborhood issues: How online communication affects offline behavior." *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*.

Falavarjani, Seyed Amin Mirlohi, et al. 2019. "The reflection of offline activities on users' online social behavior: An observational study." *Information Processing & Management* 56: 102070.

Fox, Jesse, and Margaret C. Rooney. 2015. "The Dark Triad and trait self-objectification as predictors of men's use and self-presentation behaviors on social networking sites." *Personality and Individual Differences* 76: 161-165.

Fox, Jesse, and Megan A. Vendemia. 2016. "Selective self-presentation and social comparison through photographs on social networking sites." *Cyberpsychology, Behavior, and Social Networking* 19: 593-600.

Gavai, Gaurang, et al. 2015. "Supervised and unsupervised methods to detect insider threat from enterprise social and online activity data." *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 6: 47-63.

Gibbs, Jennifer L., Nicole B. Ellison, and Rebecca D. Heino. 2006. "Self-presentation in online personals: The role of anticipated future interaction, self-disclosure, and perceived success in Internet dating." *Communication Research* 33: 152-177.

Gill, A. J., Oberlander, J., & Austin, E. 2006. "Rating e-mail personality at zero acquaintance." *Personality and Individual Differences* 40: 497-507.

Goby, V. P. 2006. “Personality and online/offline choices: MBTI profiles and favored communication modes in a Singapore study.” *CyberPsychology and Behavior* 9: 5–13.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life.* Anchor Books: New York.

Golbeck, Jennifer, Cristina Robles, and Karen Turner. 2011. "Predicting personality with social media." *CHI'11 Extended Abstracts on Human Factors in Computing Systems*.

Golbeck, Jennifer, et al. 2011. "Predicting personality from twitter." *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing.*

Gonzales, A. L., & Hancock, J. T. 2011. “Mirror, mirror on my Facebook wall: Effects of exposure to Facebook on self-esteem.” *Cyberpsychology, Behavior, and Social Networking* 14: 79–83.

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. 2011. “Manifestations of personality in online social networks: Self-reported Facebook behaviors and observable profile information.” *Cyberpsychology, Behavior and Social Networking* 14: 483-488.

Greitzer, Frank L., and Deborah A. Frincke. 2010. "Combining traditional cyber security audit data with psychosocial data: towards predictive modeling for insider threat mitigation." *Insider threats in cyber security* 2010:85-113.

Gritzalis, Dimitris, et al. 2014. "Insider threat: enhancing BPM through social media." *2014 6th International Conference on New Technologies, Mobility and Security (NTMS)*.

Guadagno, Rosanna E., Bradley M. Okdie, and Sara A. Kruse. 2012. "Dating deception: Gender, online dating, and exaggerated self-presentation." *Computers in Human Behavior* 28: 642-647.

Haferkamp, Nina, et al. 2012. "Men are from Mars, women are from Venus? Examining gender differences in self-presentation on social networking sites." *Cyberpsychology, Behavior, and Social Networking* 15: 91-98.

Hampton, K.N., Goulet, L.S., Rainie, L., & Purcell, K. 2011. “Social networking sites and our lives: How people’s trust, personal relationships, and civic and political involvement are connected to their use of social networking sites and other technologies.” *Pew Research Center*. Available at: https://www.pewresearch.org/internet/Reports/2011/Technology-and-social-networks.aspx.

Hayes, Rebecca A., Andrew Smock, and Caleb T. Carr. 2015. "Face [book] management: Self-presentation of political views on social media." *Communication Studies* 66: 549-568.

Henrich, J., S. Heine, and A. Norenzayan. 2010. “The weirdest people in the world?” *Behavioral and Brain Sciences* 33:61-83.

Heise, David, and Neil MacKinnon. 2010. *Self, identity, and social institutions*. Springer: New York.

Ho, Shuyuan Mary, and Hwajung Lee. 2012. "A Thief among Us: The Use of Finite-State Machines to Dissect Insider Threat in Cloud Communications." *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 3: 82-98.

Ho, Shuyuan Mary, and Merrill Warkentin. 2017. "Leader's dilemma game: An experimental design for cyber insider threat research." *Information Systems Frontiers* 19: 377-396.

Ho, Shuyuan Mary, et al. 2015. "Insider threat: Language-action cues in group dynamics." *Proceedings of the 2015 ACM SIGMIS Conference on Computers and People Research*.

Hocevar, Kristin Page, Andrew J. Flanagin, and Miriam J. Metzger. 2014. "Social media self-efficacy and information evaluation online." *Computers in Human Behavior* 39: 254-262.

Hogan, Bernie. 2010. "The presentation of self in the age of social media: Distinguishing performances and exhibitions online." *Bulletin of Science, Technology & Society* 30: 377-386.

Kandias, Miltiadis, et al. 2013. "Proactive insider threat detection through social media: The YouTube case." *Proceedings of the 12th ACM Workshop on Privacy in the Electronic Society*.

Kim, Hee-Woong, Hock Chuan Chan, and Atreyi Kankanhalli. 2012. "What motivates people to purchase digital items on virtual community websites? The desire for online self-presentation." *Information Systems Research* 23: 1232-1245.

Koop, Royce, and Alex Marland. 2012. "Insiders and outsiders: Presentation of self on Canadian parliamentary websites and newsletters." *Policy & Internet* 4: 112-135.

Kosinski, Michal, et al. 2014. "Manifestations of user personality in website choice and behaviour on online social networks." *Machine Learning* 95: 357.

Kozyreva, Anastasia, Stephan Lewandowsky, and Ralph Hertwig. 2020. "Citizens versus the internet: Confronting digital challenges with cognitive tools." *Psychological Science in the Public Interest* 21: 103-156.

Krämer, Nicole C., and Stephan Winter. 2008. "Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites." *Journal of Media Psychology* 20: 106-116.

Legg, Philip A., et al. 2013. "Towards a conceptual model and reasoning structure for insider threat detection." *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 4: 20-37.

Linares, Kevin, et al. 2011. "A second life within Second Life: Are virtual world users creating new selves and new lives?" *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)* 1: 50-71.

Lynch, D. M. 2006. "Securing Against Insider Attacks." *Information Security and Risk Management* 15: 39-47

Ma, Xiao, Jeff Hancock, and Mor Naaman. 2016. "Anonymity, intimacy and self-disclosure in social media." *Proceedings of the 2016 CHI conference on human factors in computing systems*.

Marcus, B., Machilek, F., & Schutz, A. 2006. "Personality in cyberspace: personal web sites as media for personality expressions and impressions." *Journal of Personality and Social Psychology* 90: 1014-1031.

Marcus, Bernd, and Heinz Schuler. 2004. "Antecedents of Counterproductive Behavior at Work: A General Perspective." *Journal of Applied Psychology* 89: 647-660.

Marriott, Tamsin C., and Tom Buchanan. 2014. "The true self online: Personality correlates of preference for self-expression online, and observer ratings of personality online and offline." *Computers in Human Behavior* 32: 171-177.

Martinez-Moyano, Ignacio J., et al. 2008. "A behavioral theory of insider-threat risks: A system dynamics approach." *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 18: 1-27.

McKenna, K. Y. A. 2007. "Through the Internet looking glass: Expressing and validating the true self." In A. Johnson, K. McKenna, T. Postmes, & U. -D. Reips (Eds.), *The Oxford Handbook of Internet Psychology* (pp. 205–221).

Mead, George Herbert. 1934. *Mind, self and society*. University of Chicago Press: Chicago.

Mehdizadeh, Soraya. 2010. "Self-presentation 2.0: Narcissism and self-esteem on Facebook." *Cyberpsychology, Behavior, and Social Networking* 13: 357-364.

Michikyan, Minas, Kaveri Subrahmanyam, and Jessica Dennis. 2014. "Can you tell who I am? Neuroticism, extraversion, and online self-presentation among young adults." *Computers in Human Behavior* 33: 179-183.

Munafò, Marcus R., et al. 2017. "A manifesto for reproducible science." *Nature human behaviour* 1: 1-9.

Nurse, Jason RC, et al. 2014. "Understanding insider threat: A framework for characterizing attacks." *2014 IEEE Security and Privacy Workshops.*

O'Brien, J. 1999. "Writing the Body: Gender Reproduction in Online Interaction," in M. Smith and P. Kollock (eds) *Communities in Cyberspace*, pp. 29–59. London: Routledge.

Ogan, Christine L., Muzaffer Ozakca, and Jacob Groshek. 2008. "Embedding the internet in the lives of college students: Online and offline behavior." *Social Science Computer Review* 26: 170-177.

Okdie, B. M., Guadagno, R. E., Bernieri, F. J., Geers, A. L., and Mclarney-Vesotski, A. R. 2011. "Getting to know you: face-to-face versus online interactions." *Computers in Human Behavior* 27: 153-159.

Ong, Eileen YL, et al. 2011. "Narcissism, extraversion and adolescents' self-presentation on Facebook." *Personality and Individual Differences* 50: 180-185.

Packer, Dominic J. 2008. "On being both with us and against us: A normative conflict model of dissent in social groups." *Personality and Social Psychology Review* 12: 50-72.

Packer, Dominic J., and Alison L. Chasteen. 2010. "Loyal deviance: Testing the normative conflict model of dissent in social groups." *Personality and Social Psychology Bulletin* 36: 5-18.

Papacharissi, Zizi. 2002. "The presentation of self in virtual life: Characteristics of personal home pages." *Journalism & Mass Communication Quarterly* 79: 643-660.

Pressman, Elaine. 2015. "Risk assessment for insider threat: critical infrastructure, military and intelligence applications." *Revista Română de Studii de Intelligence* 14: 91-116.

Randazzo, M. R., Keeney, M., Kowalski, E., Cappelli, D., and Moore, A. 2004. "Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector," U.S. Secret Service and CERT Coordination Center/Carnegie Mellon University Software Engineering Institute (http://www.secretservice.gov/ntac/its_report_040820.pdf).

Roberts, M. E. 2018. *Censored: Distraction and diversion inside China's Great Firewall*. Princeton University Press.

Ross, B. et al. 2009. "Nidal Malik Hasan, Suspected Fort Hood Shooter, Was Called "Camel Jockey" [Online]." *ABC News*. Available: http://abcnews.go.com/Blotter/nidal- malik-hasan-wanted-army-family/story?id=9008184.

Rouse, S. V. and Haas, H. A. 2003. "Exploring the accuracies and inaccuracies of personality perception following internet-mediated communication." *Journal of Research in Personality* 37: 446-467.

Rui, Jian, and Michael A. Stefanone. 2013. "Strategic self-presentation online: A cross-cultural study." *Computers in Human Behavior* 29: 110-118.

Schultz, E. Eugene. 2002. "A framework for understanding and predicting insider attacks." *Computers & security* 21: 526-531.

Schwartz, H. Andrew, et al. 2013. "Personality, gender, and age in the language of social media: The open-vocabulary approach." *PloS One* 8: e73791.

Schwartz, B. 2016. "Google's search knows about over 130 trillion pages." *Search Engine Land*. https://searchengineland.com/googles-search-indexes- hits-130-trillion-pages-documents-263378

Seepersad, Sean. 2004. "Coping with loneliness: Adolescent online and offline behavior." *CyberPsychology & Behavior* 7: 35-39.

Sibona, Christopher, and Steven Walczak. 2011. "Unfriending on Facebook: Friend request and online/offline behavior analysis." *44th Hawaii International Conference on System Sciences.*

Stryker, Sheldon, and Peter J. Burke. 2000. "The past, present, and future of an identity theory." *Social Psychology Quarterly* 63: 284-297.

Subrahmanyam, K., Reich, S. M., Waechter, N., & Espinoza, G. 2008. "Online and offline social networks: Use of social networking sites by emerging adults." *Journal of Applied Developmental Psychology* 29: 420–433.

Suler, John. 2005. "The online disinhibition effect." *International Journal of Applied Psychoanalytic Studies* 2: 184-188.

Sumner, Chris, et al. 2012. "Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets." *11th International Conference on Machine Learning and Applications. Vol. 2.*

Theoharidou, Marianthi, et al. 2005. "The insider threat to information systems and the effectiveness of ISO17799." *Computers & Security* 24: 472-484.

Toma, Catalina L., and Jeffrey T. Hancock. 2010. "Looks and Lies: The Role of Physical Attractiveness in Online Dating Self-Presentation and Deception." *Communication Research* 37: 335-351.

Toma, Catalina L., Jeffrey T. Hancock, and Nicole B. Ellison. 2008. "Separating fact from fiction: An examination of deceptive self-presentation in online dating profiles." *Personality and Social Psychology Bulletin* 34: 1023-1036.

Tosun, L. P. and Lajunen, T. 2010. "Does internet use reflect your personality? Relationship between Eysenck's personality dimensions and internet use." *Computers in Human Behavior* 26: 162-167.

Turkle, S. 1997. *Life on the Screen: Identity in the Age of the Internet*. Phoenix: London.

Vogel, Erin A., and Jason P. Rose. 2016. "Self-reflection and interpersonal connection: Making the most of self-presentation on social media." *Translational Issues in Psychological Science* 2: 294.

Vogel, Erin A., et al. 2014. "Social comparison, social media, and self-esteem." *Psychology of Popular Media Culture* 3: 206.

Wang, J. L., Jackson, L. A., Zhang, D. J., & Su, Z. Q. 2012. "The relationships among the Big Five Personality factors, self-esteem, narcissism, and sensation-seeking to Chinese University students' uses of social networking sites (SNSs)." *Computers in Human Behavior* 28: 2313–2319.

Ward, Janelle. 2016. "Swiping, matching, chatting: Self-presentation and self-disclosure on mobile dating apps." *Human IT: Journal for Information Technology Studies as a Human Science* 13: 81-95.

Widyanto, L., & Griffiths, M. D. 2011. "An empirical study of problematic Internet use and self-esteem." *International Journal of Cyber Behavior, Psychology and Learning* 1: 13–24.

Wynn, E. and Katz, J. 1997 "Hyperbole over Cyberspace: Self-presentation and Social Boundaries in internet Home Pages and Discourse," *The Information Society: an International Journal* 13: 297–328.

Yang, Chia-chen, and B. Bradford Brown. 2016. "Online self-presentation on Facebook and self-development during the college transition." *Journal of Youth and Adolescence* 45: 402-416.

Yarkoni, Tal. 2010. "Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers." *Journal of Research in Personality* 44: 363-373.

Yee, N., Bailenson, J. N., & Ducheneaut, N. 2009. "The Proteus effect: Implications of transformed digital self-representation on online and offline behavior." *Communication Research* 36: 285–312.

Zywica, J., & Danowski, J. 2008. "The faces of Facebookers: Investigating social enhancement and social compensation hypotheses: Predicting Facebook and offline popularity from sociability and self-esteem, and mapping the meanings of popularity with semantic networks." *Journal of Computer-Mediated Communication* 14: 1–34.