The dark side of social media

Guest Editors: Jari Salo, Matti Mäntymäki and A.K.M. Najmul Islam

1165 Editorial boards
1166 Guest editorial
1169 When social media traumatizes teens: the roles of online risk exposure, coping, and post-traumatic stress
Bridget Christine McHugh, Pamela Wisniewski, Mary Beth Rosson and John M. Carroll
1189 Coping with mobile technology overload in the workplace
Pengzhen Yin, Carol X.J. Ou, Robert M. Davison and Jie Wu
1213 Is my fear of missing out (FOMO) causing fatigue? Advertising, social media fatigue, and the implications for consumers and brands
Laura Frances Bright and Kelty Logan
1228 Investigating microblogging addiction tendency through the lens of uses and gratifications theory
Qian Li, Xunhui Guo, Xue Bai and Wei Xu
1253 Social media’s have-nots: an era of social disenfranchisement
Xinru Page, Pamela Wisniewski, Bart P. Krynenburg and Moses Namara
1275 Tolerating and managing extreme speech on social media
Brett G. Johnson
1292 The dark side of news community forums: opinion manipulation trolls
Todor Mihaylov, Tsvetomila Mihaylova, Preslav Nakov, Lluís Màrquez, Georgi D. Georgiev and Ivan Kolev Koychev
1313 Online moral disengagement and hostile emotions in discussions on hosting immigrants
Francesca D’Errico and Mannella Paciello
1336 The bidirectional mistrust: callers’ online discussions about their experiences of using the national telephone advice service
Annica Björkman and Martin Salzmann-Erikson

ISBN 978-1-78769-903-8

www.emeraldinsight.com/loi/intr
The dark side of social media – and *Fifty Shades of Grey* introduction to the special issue: the dark side of social media

The massive-scale diffusion of social media in private life as well as workplaces calls for increased awareness and understanding of its risks and adverse consequences for individuals, organizations and societies (Garcia and Sikström, 2014; Fox and Moreland, 2015; Mäntymäki and Islam, 2016; Oberst *et al.*, 2017; Trekels *et al.*, 2018). For example, the recent discussion on the role of social media in amplifying socially undesirable phenomena such as fake news (Allcott and Gentzkow, 2017) and hate speech (Farkas *et al.*, Forthcoming) demonstrates a clear need for a critical analysis of the undesirable societal consequences of social media. At an individual level, the influence of people’s social media engagement on phenomena such as loneliness (Song *et al.*, 2014), depression (Lin *et al.*, 2016), envy (Tandoc *et al.*, 2015) and narcissism (Andreassen *et al.*, 2017) is everything but perfectly understood. Finally, there is a long-standing void in the literature with respect to paradoxes and controversies related to the nature and outcomes of the core uses of social media. For example, following other people on social networking sites can be a means to maintain interpersonal connections but may also turn into digital voyeurism (Mäntymäki and Islam, 2014; Mäntymäki and Islam, 2016).

With respect to future research directions regarding the dark side of social media, we highlight three areas that in our opinion would deserve increased scholarly attention. First, there is a clear need for research focusing on the dark side of social media particularly in the workplace context. Different social media tools and applications have rapidly proliferated to workplaces (Mäntymäki and Riemer, 2016). However, the extant literature has largely focused on social media use in people’s private lives. As a result, future research could for example investigate the relationship between workplace social media use and information overload (Yin *et al.*, 2018) and technostress.

Second, while the literature has identified a number of psycho-social phenomena attributed to social media use, there is a void in the current understanding of how the characteristics social media platforms and users’ activities may contribute to the negative individual-level consequences. As a result, increased attention to the IT artefact (Mäntymäki and Brown, 2014) could be a means to make the research more relevant for service designers and platform operators.

Third, to take into account the potential social desirability bias (Fisher, 1993) related to people’s responses and perceptions regarding their own cyberbullying, narcissism, trolling or voyeurism, future research should more extensively adopt research designs that utilize data from multiple sources. Hence, we recommend increased use of mixed methods (Venkatesh *et al.*, 2013) and multi-method (Mäntymäki and Riemer, 2014) research designs as well as using data about users’ actual behaviors to mitigate the risk of social desirability bias.

The purpose of this *Internet Research* Special Issue “The Dark Side of Social Media” is to provide a venue for critical academic research focusing on the adverse aspects of social media. The large number and diversity of themes covered by papers submitted to the special issue demonstrate that scholars from a range of disciplinary backgrounds and methodological premises are investigating the phenomena surrounding social media from a critical perspective. The topics of the submissions ranged from extreme manifestations of the dark side of social media, such as pedophilia, hate crime and terrorism, to more mundane phenomena, such as fear of missing out (FOMO) and information overload. This exemplifies
that the dark side of social media in fact comprises a number of different degrees of
darkness – and shades of gray.

With the invaluable help and support of the reviewers, we have selected eight papers for
inclusion in the special issue, representing the two main lines of inquiry from the received
submissions, negative aspects of social media for individuals and analyzing large volumes
of online discussion data.

The first paper, “When social media traumatizes teens: the roles of online risk exposure,
coping, and post-traumatic stress,” examines online risk experiences, particularly
information breaches, exposure to explicit content, cyberbullying and sexual solicitation,
as causes of post-traumatic stress disorder among teens. The second paper, “Coping with
mobile technology overload in the workplace,” investigates mobile information technology
overload in a workplace context. The third paper, “Is my FOMO causing fatigue? Advertising,
social media fatigue, and the implications for consumers and brands,” examines how advertising factors such as attitude and intrusiveness impact social media
fatigue, fear of missing out (FOMO) and privacy interlink. The fourth paper, “Investigating
microblogging addiction tendency through the lens of use and gratification,” examines the
behavioral and perceptual factors that impact the addiction tendency levels of
microblogging users. The fifth paper, entitled “Social media’s have-nots: an era of social
disenfranchisement,” focuses on the motivations, concerns, benefits and consequences
associated with non-use of social media.

The sixth paper, “Tolerating and managing extreme speech on social media,” seeks to
understand how internet users tolerate and manage extreme speech. The seventh paper,
“The dark side of news community forums: opinion manipulation trolls,” explores the dark
side of news community forums by identifying suspicious users and regular users. The eighth
paper, “Online moral disengagement and hostile emotions in discussions on hosting
immigrants,” seeks to examine the relationship between moral disengagement mechanisms
and the hostile emotions expressed online towards hosting immigrants. The last paper,
“The bidirectional mistrust – Callers’ online discussions about their experiences of using
the national telephone advice service,” discloses the dark side of internet-based forums in which
callers discuss their experiences of using the national telephone advice service.

Jari Salo
Department of Economics and Management, Marketing Unit,
Faculty of Agriculture and Forestry, University of Helsinki, Helsinki, Finland, and
Matti Mäntymäki and A.K.M. Najmul Islam
University of Turku, Turku, Finland

References
Economic Perspectives, Vol. 31 No. 2, pp. 211-236.

social media, narcissism, and self-esteem: findings from a large national survey”, Addictive
Behaviors, Vol. 64, January, pp. 287-293.

propaganda in social media”, New Media & Society, available at: https://doi.org/10.1177/
1461444817707759


relational and psychological stressors associated with Facebook use and affordances”,


When social media traumatizes teens
The roles of online risk exposure, coping, and post-traumatic stress

Bridget Christine McHugh
Department of Psychology, University of Central Florida, Orlando, Florida, USA

Pamela Wisniewski
Department of Computer Science, University of Central Florida, Orlando, Florida, USA, and

Mary Beth Rosson and John M. Carroll
College of Information Sciences and Technology, Pennsylvania State University, University Park, Pennsylvania, USA

Abstract
Purpose – The purpose of this paper is to examine the extent to which negative online risk experiences (information breaches, explicit content exposure, cyberbullying and sexual solicitations) cause post-traumatic stress disorder (PTSD) symptoms in adolescents. The study also explores whether teens’ short-term coping responses serve to mitigate PTSD or, instead, act as a response to stress from online events.

Design/methodology/approach – The study utilized a web-based diary design over the course of two months. Data were analyzed using hierarchical linear modeling with repeated measures.

Findings – The study confirmed that explicit content exposure, cyberbullying and sexual solicitations (but not information breaches) evoke symptoms of PTSD. Analyses also indicated that teens engage in active and communicative coping after they experience post-traumatic stress, regardless of risk type or frequency.

Practical implications – The authors found that teens took active measures to cope with online risks soon after they felt threatened (within a week). Actively coping with stressful situations has been shown to enhance adolescent resilience and reduce long-term negative effects of risk exposure. If these early coping behaviors can be detected, social media platforms may be able to embed effective interventions to support healthy coping processes that can further protect teens against long-term harm from exposure to online risks.

Originality/value – This is the first study to examine situational PTSD symptoms related to four types of adolescent online risk exposure within the week exposure occurred. By applying two competing theoretical frameworks (the adolescent resilience framework and transactional theory of stress), the authors show empirical evidence that suggests short-term coping responses are likely a stress reaction to PTSD, not a protective factor against it.

Keywords Coping, Post-traumatic stress disorder, Cyberbullying, Online privacy, Online safety, Online sexual solicitation

Paper type Research paper

1. Introduction
Most adolescents use social media daily (Lenhart et al., 2010), which necessitates an examination of the potential “dark side” of social media use for teens. Adolescent internet use has been associated with decreased well-being (Kraut et al., 1998), and excessive use has been tied to depression (Young and Rogers, 1998), anxiety (Dalbudak et al., 2013), aggression (Lim et al., 2015), and social isolation (Kraut et al., 1998). Teens who use social media and similar platforms excessively may also develop addictive behaviors (Balakrishnan and...
Shamim, 2013), especially when they have fewer offline social ties (Yang et al., 2016) and suffer from social anxiety (Elhai et al., 2018). These addictive behaviors often lead to poorer mental health (Xue et al., 2018) and an increased risk of identity theft via malicious profiles (Rose, 2011). Teens may also use social media to fulfill adverse gratifications such as voyeurism and exhibitionism (Mantymäki and Islam, 2016). Negative effects of social media on mood have been related to the nature of teens’ online interactions and viewing behavior (Lee et al., 2015), suggesting that exposure to certain online risks may be detrimental to teens’ developmental growth (Burk et al., 2014). This has led to concern over the emotional and psychological effects of risks teens encounter online, including sexual solicitations (Rice et al., 2015), privacy breaches (Berriman and Thomson, 2015; Berry, 2004), cyberbullying (O’Keeffe and Clarke-Pearson, 2011), and explicit content exposure (van Oosten, 2015). Although parents may restrict access to certain websites, this may still not shield them from online risks (Peters, 2006), as many risks teens encounter online occur on social networking sites (Mitchell et al., 2014). Given the pervasiveness of social media use among teens (Forsyth et al., 2013), it is unrealistic to completely prevent online risk exposure. Thus, some researchers have shifted away from restricting teens’ internet behaviors to addressing the risks that teens may inevitably encounter online, so that they can be resilient against them (Wisniewski et al., 2015).

Our study builds on this resilience-based perspective by conducting an “in-situ” two-month long diary study, which asked 75 teens to report their online risk experiences the week they occurred. This paper makes novel contributions to the adolescent online safety and risk literature in the following ways. First, we measure teens’ episodic online risk experiences over a two-month period, as opposed to traditionally used cross-sectional approaches. Second, our diary prompts differentiated between four distinct types of online risks: information breaches, explicit content exposure, cyberbullying and sexual solicitations. Broadening our definition of online risks allowed us to compare across the four risk types and identify distinct differences regarding their effects. Third, when a teen reported experiencing an online risk event, we asked follow-up questions regarding how teens coped with each experience. Previous research has examined risk factors that lead to online risk exposure (Dredge et al., 2014), but our study examines teens’ coping behaviors in direct response to a particular risk event. Thus, the present study focuses on how teens’ own actions rather than their circumstances shape their experience of online risk exposure. We are also the first to utilize a pre-validated psychological measure (the Child’s Revised Impact of Event Scale or CRIES; Perrin et al., 2005) for measuring clinical post-traumatic stress disorder (PTSD) symptoms across all four online risk categories of episodic online risk occurrences.

Combined with the methodological rigor used in our study design, we also make significant contributions to theory by applying two competing theoretical models to our empirical data set, the resilience framework (Fergus and Zimmerman, 2005) and the transactional theory of stress (Lazarus, 1966), to show how coping responses exhibited soon after an online risk event are likely reactions induced by PTSD, not protective factors against PTSD. Finally, we discuss practical implications for information online safety policies and laws, as well as design implications for social media developers to help teens more effectively cope with the risks.

2. Background: the dark-side of social media for teens
Our research examines the potential dark side of social media and online engagement for teens. Recent research has begun to illuminate the potential psychological effects of social media use, such as a detrimental impact on adolescent self-esteem (Kross et al., 2013), and increased risk for depression (O’Keeffe and Clarke-Pearson, 2011). Social media sites encourage users to present the most positive aspects of their lives (Krämer and Winter, 2008), leading to dissatisfaction via social comparisons (Coyne et al., 2017) and “Facebook
depression” (O’Keeffe and Clarke-Pearson, 2011). Adolescents may even develop social media addiction after excessive use (Balakrishnan and Shamim, 2013). While such psychological effects of social media use may be fairly insidious, social media can also directly expose teens to a myriad of risky online situations. In the next section, we provide a review of the adolescent online safety and risk literature to delineate these online risks.

2.1 A labyrinth of online risks
Drawing from the seminal work of Livingstone and Smith (2014), who differentiated between harmful content (e.g. explicit content exposure), contact (e.g. sexual solicitation and cyberbullying), and conduct (e.g. information breaches), we identified four primary categories of online risk from the literature: information privacy breaches, explicit content exposure, cyberbullying and sexual solicitations.

Though teens express concern about their privacy, they engage in behaviors that put them at risk for information privacy breaches (Barnes, 2006). Privacy concerns have little to no impact on adoption (Tan et al., 2012) and self-disclosure (Cheung et al., 2015) on social media sites. Privacy breaches often occur because of features included in social media sites, such as tagging other users without prior consent (Birnholtz et al., 2017), automated geotagging (Albrecht and McIntyre, 2015) or other location-based features that may cause privacy concerns (Zhou, 2017). Social media also relies on users to self-report explicit content exposure and often has delayed and inconsistent enforcement of content policies (Crawford and Gillespie, 2016). Thus, teens may be exposed to images or videos that are overly violent (e.g. wars; De Choudhury et al., 2014), or content that contain self-harm and other immoral and illicit behaviors (Wisniewski et al., 2016) that may be disturbing to young viewers (Boyd and Swanson, 2016). Meanwhile, the anonymity of social media may also put teens at risk for cyberbullying, while protecting bullies (Barlett et al., 2016), and increasing the sense of fear and powerlessness of victims (Dooley et al., 2009). There is also concern about sexual solicitations teens may receive from other social media users, such as peers asking for nude photos. Indeed, 7 percent of teens indicate they have sent a nude photo to someone (Ybarra and Mitchell, 2014). As many social media users are much older (Bogdanova et al., 2012), these sites also put teens at higher risk of sexual predation (Cano et al., 2014).

2.2 Gaps in the literature
Past studies examining adolescent online risks have often taken case-based (McCarty et al., 2011) or cross-sectional approaches that try to understand risk prevalence at a population level (Mitchell et al., 2011). Large-scale studies often ignore the context of risks, only collecting dichotomous “yes” and “no” responses as to whether or not teens experienced online risks “ever” (Gross, 2004). Further, with cross-sectional designs, data about risks are collected long after the incident occurred (the best case is usually “within the past year”; Jones et al., 2013; Smith et al., 2014). A common theme among the literature is that the research typically focuses on risk exposure, or one type of risk (typically cyberbullying) in isolation from all others, as opposed to what occurs after exposure (Pinter et al., 2017).

Some recent research has focused on how teens cope after experiencing stressful events online (e.g. blocking cyberbullies; Orel et al., 2015), especially on why teens select specific coping strategies for certain online risks, and how employing these coping strategies may increase resilience (Raskauskas and Huyhn, 2015). Yet, studies on risk-coping tend to use more qualitative approaches, such as asking teens to self-report their emotional reactions (e.g. Nie and Erbring, 2000). To our knowledge, no other empirical studies have been conducted over time to measure all four types of episodic online risk experiences teens have on a weekly basis, nor have they systematically examined coping behaviors or PTSD in relation to these risks. Therefore, the novelty of our approach sets our research apart from what has been done in the past.
3. Application of theory and research framework

3.1 Risk and adolescent resilience

Our work was initially motivated by Fergus and Zimmerman’s (2005) framework for adolescent resilience. Resiliency is the ability to recover from an emotional trauma (Lazarus, 1966). Teens who are more resilient experience less severe and fewer emotions after an event (Kobasa, 1979). Resilience can be strengthened over time (McAllister and McKinnon, 2009). Depending on these risk and protective factors, teens may be less resilient, and therefore more likely to suffer negative outcomes of risk exposure including depression (Erdem and Slesnick, 2010) and delinquency (Glowacz and Born, 2015). Most previous research has focused on exposure itself and risk factors, not on the extent to which these risks actually cause emotional harm (Slavtcheva-Petkova et al., 2015; Livingstone and Smith, 2014) and tends to assume all risk exposure is harmful (Pinter et al., 2017). In contrast, our work specifically examines the relationship between online risk exposure and post-exposure symptoms of PTSD. Further, we apply two competing but relevant theories to better understand the role coping plays in this process. Our research framework is illustrated in Figure 1, and the constructs of our model are described below.

3.2 Post-traumatic stress disorder

To differentiate between online risk exposure and harm, our work supplements resilience theory with research on PTSD. PTSD is a clinically diagnosable condition that arises from exposure to negative events. Unlike traumatization symptoms that are general and somatic (Briere and Elliott, 2003), PTSD is event-specific (Green et al., 1985). Risk type and individual factors influence the likelihood that PTSD will occur (Ozer et al., 2008). Symptoms include avoidance of reminders of a specific event, hyper-arousal in similar situations and intrusive thoughts about the traumatic event (Perrin et al., 2005). PTSD has been associated with unwanted sexual solicitations (Fitzgerald et al., 1997), bullying (Spence Laschinger and Nosko, 2015), and explicit content exposure (Clohessy and Ehlers, 1999) in offline contexts. While some research has examined post-traumatic stress in online contexts, this research examined severe trauma related to cyberbullying (Ranney et al., 2016) and online sexual exploitations (Wells and Mitchell, 2007) in specific high-risk contexts. For instance, Ranney et al. (2016) surveyed youth who sought out emergency medical services, while Wells and Mitchell (2007) surveyed mental health professionals who reported on patients who were being treated for problematic internet experiences. Other research examined PTSD

![Research framework](image-url)
symptoms resulting from cyberbullying in high-risk youth populations, such as within the LGBT community (Beckerman and Auerbach, 2014) and teens at high risk of psychosis (Magaud et al., 2013). Similarly, others who have studied online risks more generally tend to focus on risk prevalence (Mitchell et al., 2011), or on severe online risks (e.g. cyber-sexual assault; Holladay, 2016), not risks typical teens encounter on a weekly basis. Since the association between PTSD and the frequency and type of risks teens encounter online is unclear, we must first examine whether these events are traumatic enough to warrant the need for resilience. Therefore, our first research question is:

**RQ1.** Is teen exposure to online risks associated with symptoms of PTSD?

### 3.3 Adolescent risk-coping

The resilience framework also suggests that the negative outcomes of traumatic events may be mitigated by various protective factors. For instance, teens who have supportive mentoring relationships (Hurd and Zimmerman, 2010) experience significantly less stress and a quicker recovery period than other teens exposed to the same situation (Fergus and Zimmerman, 2005). Thus, teens with protective resources have better outcomes following a negative event (Cline et al., 2014), as well as lower anxiety and depression (Anyan and Hjemdal, 2016) than other teens who have been exposed to the same risks. Protective factors are usually framed as external to the individual (e.g. social support; Fergus and Zimmerman, 2005). However, resilience theory also posits that teens’ own internal assets, or personal traits, can promote resiliency and may also protect against the negative effects of risk exposure. While the effect of teens’ external resources on outcomes of risk exposure have been studied in the past (Hinduja and Patchin, 2008), research on teens’ assets have rarely been applied to online risk exposure, and only to certain online risks (Orel et al., 2015; Raskauskas and Huyhn, 2015).

Coping is one of the most commonly studied assets exhibited by teens (Wills et al., 1996). Coping behaviors can be active (i.e. actions to remove a stressor; Carver et al., 1989), passive (i.e. avoiding a stressor; Connor-Smith and Flachsbart, 2007), or communicative (i.e. talking about the stressor; Saunders et al., 2016). Coping has a strong protective effect against PTSD symptoms (Clohessy and Ehlers, 1999). While the protective effects of coping are well established (Wills et al., 1996), there is little research on the protective effects of adolescent risk-coping behaviors within online contexts. Since the relationship between resilience and online risk-coping is under-studied, we ask the following research question:

**RQ2.** (Resilience theory): Do coping behaviors moderate the relationship between online risk exposure and PTSD symptoms?

Note that in our case, **RQ1** is a prerequisite for answering **RQ2** because a moderating relationship implies that there is a significant, direct relationship between online risk exposure and PTSD that requires mitigation from online coping.

### 3.4 Transactional theory of stress as an alternative

There is a lack of clarity in the literature as to whether coping is primarily a protective asset that insulates teens from harm resulting from risk exposure (Wills et al., 1996) or if coping is primarily a response to stress from a situation (Lazarus, 1966). To empirically test these competing theories, we also frame coping using the transactional theory of stress (Lazarus, 1966), which takes a broader view of the antecedents and outcomes of negative life events. While the resilience framework views risk in terms of antecedents (assets and resources), moderation effects (protective factors), and negative outcomes (e.g. PTSD), the transactional theory of stress (Lazarus, 1966) suggests multiple stages of risk assessment marked by appraisals of the event. In the primary appraisal stage, when the victim is experiencing a
stressful event, the individual determines the potential for harm, while experiencing their initial emotional and behavioral reaction. In the latter stage, individuals form a secondary appraisal, evaluating their stress levels and responding with coping behaviors (Folkman et al., 1986). Thus, the transactional theory of stress (Lazarus, 1966) proposes that coping may occur in response to negative emotional outcomes. As such, coping may be a response to PTSD, as documented in offline contexts (Coyne and Lazarus, 1980). Since the cross-sectional nature of previous online risk literature has not allowed for a systematic examination of coping behaviors in direct response to a specific risk event, an alternative research question examines the role of coping as a reaction to PTSD:

RQ3. (Transactional theory of stress): Are coping behaviors exhibited soon after online risk exposure a direct response to (outcome of) symptoms of PTSD?

4. Methods
4.1 Diary study overview
We conducted a two-month web-based diary study of 75 teens (ages 13-17) who reported their online risk experiences each week. We used event contingent diary methods because this method yields more accurate information than retrospective self-report data in cross-sectional studies. Cross-sectional studies are susceptible to recall error that leads to inaccurate reports when collected long after an event has occurred, mainly because individuals are more likely to recall only the most salient (or traumatic) experiences that happened to them over the course of a year instead of more frequent but less memorable experiences (Gorin and Stone, 2001).

Teens were given a unique login to an online “Diary Dashboard” where they could view their past diary entries, as well as complete new diary entries over a rolling, eight-week period. Teens were reminded to complete or finish weekly diary entries via e-mail. Parental consent was required but parents were not given access to their child’s diary portal to protect the teen’s privacy. To account for whether communication with parents affected study outcomes, the measures for communicative coping (described below) included a question asking if teens spoke with their parents or a trusted adult. In general, communication between teens and parents during the study was low (Table I).

4.2 Measures
Table I provides the psychometric properties and descriptive statistics for each construct in our model, along with each construct’s definitions.

4.2.1 Teen online risk exposure. Each weekly diary entry included questions related to four online risk categories: information breaches; cyberbullying; sexual solicitations and explicit content exposure. The frequency of online risk exposure was reported using a five-point Likert scale (1 = never in that week, 5 = six or more times that week). Participants could report no risks, a single risk type, or multiple risk types each week. It is important to note that risk experiences were measured across all online platforms, not tied to a specific social media site (e.g. Facebook). This design choice makes our results more generalizable to the true social ecologies’ perspective of teen multi-platform use (Zhao et al., 2016) than if we had tied our study to one social media platform.

4.2.2 Teen coping behaviors. Participants indicated whether or not (e.g. yes or no) they engaged in various coping behaviors after reporting an online risk experience (Table I). The behaviors were based on items from previous surveys (Jia et al., 2015; Livingstone et al., 2010; Wisniewski et al., 2016). These studies unfortunately did not establish construct validity for the scales. To determine the structure of our coping checklist, exploratory factor analysis was used. Three factors aligned with risk-coping theories of active coping
Post-traumatic stress disorder (PTSD) symptoms

- Intrusion – unwanted thoughts, images, and feelings related to the event (Horowitz et al., 1979)  
  \[ \alpha = 0.75, M = 1.82, SD = 0.89 \]
- Avoidance – active efforts to avoid thoughts or reminders of the event (Horowitz et al., 1979)  
  \[ \alpha = 0.76, M = 1.96, SD = 0.91 \]
- Arousal – heightened physiological arousal (Horowitz et al., 1979)  
  \[ \alpha = 0.71, M = 1.59, SD = 0.72 \]

Online risk exposure

- Information breaches – unwanted sharing of information or photos (Wisniewski et al., 2016)  
  \[ \alpha = 0.70, M = 1.24, SD = 0.30 \]
- Cyberbullying – deliberate, threatening or embarrassing online interactions (Wisniewski et al., 2016)  
  \[ \alpha = 0.97, M = 1.29, SD = 0.49 \]
- Sexual Solicitations – sexual interactions or requests (Wisniewski et al., 2016)  
  \[ \alpha = 0.72, M = 1.25, SD = 0.48 \]
- Explicit content exposure – voluntary or accidental viewing of pornographic, extremely violent, or deviant online content (Wisniewski et al., 2016)  
  \[ \alpha = 0.69, M = 1.24, SD = 0.30 \]

Coping behaviors

- Passive coping – behaviors that deny or ignore the stressor (Carver et al., 1989)  
  \[ \alpha = 0.63, M = 0.55, SD = 0.40 \]
- Active coping – behaviors that attempt to remove the stressor (Connor-Smith and Flachsbart, 2007)  
  \[ \alpha = 0.73, M = 0.30, SD = 0.33 \]
- Communicative coping – communicating about the stressor (Saunders et al., 2016)  
  \[ \alpha = 0.65, M = 0.26, SD = 0.32 \]

Notes: \( N = 222 \). Information breaches, cyberbullying, sexual solicitations and exposure to explicit content items were measured on a five-point Likert Scale ranging from 1 = not at all that week, 2 = once that week, 3 = 2–3 times that week, 4 = 4–5 times that week, 5 = 6 or more times in that week. Coping items were measured using a yes/no response option with 0 = no and 1 = yes. A value of 1 on the online risk exposure scale indicates no risk events occurred. KR-20 = Kuder-Richardson 20; \( \alpha = \) Cronbach’s \( \alpha \).
over the last year, was much higher for passive coping (0.85), active coping (0.83), and communicative coping (0.91).

4.2.3 Post-traumatic stress disorder symptoms. Each time a teen reported that they had been exposed to an online risk, PTSD symptoms were measured using the CRIES-13 (Perrin et al., 2005). The CRIES-13 is a pre-validated measure of three types of clinical PTSD symptoms: arousal (e.g. hypervigilance), intrusion (e.g. persistent, unwanted thoughts about events), and avoidance (e.g. avoiding reminders of events) in relation to a particular adverse event (Giannopoulou et al., 2006).

4.3 Data analysis approach
Analysis of repeated measures data must account for between-person’s variance (i.e. portion of variations in PTSD due to differences between teens) and within-person’s variance (i.e. portion of variations in PTSD for the same teen across weeks). Most of the variance in each dependent variable originated from within-person differences, as suggested by case one of the intraclass correlation coefficients (ICC (1)) calculated from the variance components (intrusion ICC (1) = 0.49, avoidance ICC (1) = 0.06, arousal ICC (1) = 0.01). As random coefficients modeling controls for within-person differences (Longford, 1994), we utilized this statistical technique to explore each research question through a series of models (Gräsbeck and Fellman, 1968). Following recommendations for random coefficients models, independent variables without a meaningful zero were grand mean centered (Hofmann and Gavin, 1998). We utilized SAS Enterprise 64’s mixed procedure to calculate beta weights ($\beta$) and determined their significance using $p$ values for two-tailed tests of significance, with a cutoff of 0.05 significance and 0.10 for marginal significance.

To answer RQ1, the effects of weekly online risk exposure on arousal, avoidance, and intrusion symptoms were examined in separate models. The week was included as a covariate to control for teens who reported fewer risks over time. RQ2 was examined by modeling the moderating effects of each coping type onto each PTSD dimension. Moderation was calculated by creating an interaction term. These interaction variables were generated by creating a new variable that consisted of the product of the two primary variables (i.e., frequency of online risk exposure and frequency of coping behaviors; see Table II) as recommended by the literature (Dalal and Zickar, 2012). For RQ3, we modeled each PTSD symptom as an independent variable with each type of coping behavior as a dependent variable.

4.4 Recruitment and sample profile
Teens were recruited through over 700 organizations that served youth and through a contact list maintained by the university. Both teens and their parents consented to participate in the study. Participants were told they would receive a $25 gift card for completing the pre-survey and up to $50 on a gift card for completing all weekly diary entries and the post-survey. 98 teens registered and 75 completed the study. Since data collection was virtual, participants were not in a single location. Though most participants (74 percent) were from Pennsylvania, they were from different regions of Pennsylvania. The remaining participants were spread across twelve other states. Ages ranged between 13 and 17 ($M = 14.79$, $SD = 1.30$). The majority were 14 (31 percent), followed by 15 (21 percent), 13 (17 percent), 16 (17 percent), and 17 (13 percent). Participants were predominately female (63 percent) and Caucasian (73 percent; 13 percent African-American, 5 percent Hispanic, 3 percent Asian, and 5 percent Other). Only 1 percent of participants indicated that they did not go online every day or almost every day. Most teens (60 percent) were from two-parent households and many (56 percent) were from households with an income of $60,000 or more.
Intrusion Arousal Avoidance

$\Delta D$ $\beta$ SE $t$ $\Delta D$ $\beta$ SE $t$ $\Delta D$ $\beta$ SE $t$

### Main effects of online risk exposure

<table>
<thead>
<tr>
<th>Risk Type</th>
<th>$\Delta D$</th>
<th>$\beta$</th>
<th>SE</th>
<th>t</th>
<th>$\Delta D$</th>
<th>$\beta$</th>
<th>SE</th>
<th>t</th>
<th>$\Delta D$</th>
<th>$\beta$</th>
<th>SE</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit content exposure</td>
<td>14.10a</td>
<td>0.11</td>
<td>0.10</td>
<td>1.20</td>
<td>11.20a</td>
<td>0.03</td>
<td>0.07</td>
<td>1.90**</td>
<td>12.40a</td>
<td>0.32</td>
<td>0.19</td>
<td>1.89*</td>
</tr>
<tr>
<td>Sexual solicitation</td>
<td>14.10a</td>
<td>0.35</td>
<td>0.14</td>
<td>2.52***</td>
<td>11.20a</td>
<td>-0.02</td>
<td>0.11</td>
<td>-0.67</td>
<td>12.40a</td>
<td>0.20</td>
<td>0.14</td>
<td>1.43</td>
</tr>
<tr>
<td>Cyberbullying</td>
<td>14.10a</td>
<td>0.20</td>
<td>0.19</td>
<td>1.20</td>
<td>11.20a</td>
<td>0.33</td>
<td>0.14</td>
<td>2.37**</td>
<td>12.40a</td>
<td>0.33</td>
<td>0.14</td>
<td>2.37**</td>
</tr>
<tr>
<td>Information breach</td>
<td>14.10a</td>
<td>-0.24</td>
<td>0.32</td>
<td>-0.92</td>
<td>11.20a</td>
<td>-0.01</td>
<td>0.24</td>
<td>-0.24</td>
<td>12.40a</td>
<td>-0.19</td>
<td>0.31</td>
<td>-0.71</td>
</tr>
</tbody>
</table>

### Main effects of coping behaviors

<table>
<thead>
<tr>
<th>Coping Type</th>
<th>$\Delta D$</th>
<th>$\beta$</th>
<th>SE</th>
<th>t</th>
<th>$\Delta D$</th>
<th>$\beta$</th>
<th>SE</th>
<th>t</th>
<th>$\Delta D$</th>
<th>$\beta$</th>
<th>SE</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active coping</td>
<td>16.40a</td>
<td>0.74</td>
<td>0.23</td>
<td>3.24***</td>
<td>26.30a</td>
<td>0.10</td>
<td>0.04</td>
<td>2.30***</td>
<td>20.80a</td>
<td>1.03</td>
<td>0.21</td>
<td>4.83***</td>
</tr>
<tr>
<td>Passive coping</td>
<td>16.40a</td>
<td>0.01</td>
<td>0.16</td>
<td>0.06</td>
<td>26.30a</td>
<td>-0.02</td>
<td>0.03</td>
<td>-0.70</td>
<td>20.80a</td>
<td>0.01</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td>Communicative coping</td>
<td>16.40a</td>
<td>0.28</td>
<td>0.23</td>
<td>1.20***</td>
<td>26.30a</td>
<td>0.12</td>
<td>0.04</td>
<td>2.72***</td>
<td>20.80a</td>
<td>0.29</td>
<td>0.22</td>
<td>1.31</td>
</tr>
</tbody>
</table>

### Interaction effects

<table>
<thead>
<tr>
<th>Coping Interaction</th>
<th>$\Delta D$</th>
<th>$\beta$</th>
<th>SE</th>
<th>t</th>
<th>$\Delta D$</th>
<th>$\beta$</th>
<th>SE</th>
<th>t</th>
<th>$\Delta D$</th>
<th>$\beta$</th>
<th>SE</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active coping interactions</td>
<td>9.0a</td>
<td>0.37</td>
<td>0.20</td>
<td>1.89</td>
<td>9.5a</td>
<td>0.32</td>
<td>0.15</td>
<td>2.17**</td>
<td>35.7a</td>
<td>0.37</td>
<td>0.19</td>
<td>1.96**</td>
</tr>
<tr>
<td>Sexual solicitation</td>
<td>29.5a</td>
<td>0.91</td>
<td>0.31</td>
<td>2.92***</td>
<td>9.5a</td>
<td>0.60</td>
<td>0.25</td>
<td>2.40**</td>
<td>35.4a</td>
<td>0.83</td>
<td>0.32</td>
<td>2.61**</td>
</tr>
<tr>
<td>Cyberbullying</td>
<td>15.2a</td>
<td>0.18</td>
<td>0.40</td>
<td>0.46</td>
<td>11.1a</td>
<td>0.56</td>
<td>0.30</td>
<td>1.87</td>
<td>28.4a</td>
<td>0.47</td>
<td>0.38</td>
<td>1.23</td>
</tr>
<tr>
<td>Information breach</td>
<td>17.7a</td>
<td>0.16</td>
<td>0.62</td>
<td>0.25</td>
<td>3.8a</td>
<td>0.05</td>
<td>0.50</td>
<td>0.10</td>
<td>32.9a</td>
<td>0.21</td>
<td>0.60</td>
<td>0.35</td>
</tr>
<tr>
<td>Passive coping interactions</td>
<td>3.1</td>
<td>-0.04</td>
<td>0.18</td>
<td>-0.24</td>
<td>10.1a</td>
<td>-0.11</td>
<td>0.14</td>
<td>-0.83</td>
<td>9.70a</td>
<td>-0.19</td>
<td>0.18</td>
<td>-1.06</td>
</tr>
<tr>
<td>Sexual solicitation</td>
<td>12.5a</td>
<td>-0.33</td>
<td>0.26</td>
<td>-1.28</td>
<td>8.0a</td>
<td>-0.46</td>
<td>0.20</td>
<td>-2.28**</td>
<td>9.70a</td>
<td>-0.57</td>
<td>0.27</td>
<td>-2.17**</td>
</tr>
<tr>
<td>Cyberbullying</td>
<td>4.0</td>
<td>-0.51</td>
<td>0.33</td>
<td>-1.54</td>
<td>12.3a</td>
<td>-0.32</td>
<td>0.25</td>
<td>-1.30</td>
<td>4.0</td>
<td>-0.27</td>
<td>0.33</td>
<td>-0.82</td>
</tr>
<tr>
<td>Information breach</td>
<td>9.0a</td>
<td>-0.77</td>
<td>0.52</td>
<td>-1.49</td>
<td>4.5</td>
<td>-0.16</td>
<td>0.40</td>
<td>-0.39</td>
<td>11.1a</td>
<td>0.56</td>
<td>0.53</td>
<td>-1.07</td>
</tr>
<tr>
<td>Communicative coping interactions</td>
<td>10.9a</td>
<td>0.31</td>
<td>0.23</td>
<td>1.35</td>
<td>23.1a</td>
<td>0.22</td>
<td>0.17</td>
<td>1.31</td>
<td>18.4a</td>
<td>0.25</td>
<td>0.22</td>
<td>1.15</td>
</tr>
<tr>
<td>Sexual solicitation</td>
<td>25.8a</td>
<td>1.24</td>
<td>0.38</td>
<td>3.29**</td>
<td>18.4a</td>
<td>0.72</td>
<td>0.30</td>
<td>2.38**</td>
<td>21.5a</td>
<td>1.19</td>
<td>0.40</td>
<td>3.03**</td>
</tr>
<tr>
<td>Cyberbullying</td>
<td>10.6a</td>
<td>0.45</td>
<td>0.41</td>
<td>1.09</td>
<td>16.9a</td>
<td>0.46</td>
<td>0.30</td>
<td>1.52</td>
<td>11.2a</td>
<td>0.27</td>
<td>0.41</td>
<td>0.65</td>
</tr>
<tr>
<td>Information breach</td>
<td>12.0a</td>
<td>1.38</td>
<td>0.71</td>
<td>2.11a</td>
<td>24.1a</td>
<td>1.03</td>
<td>0.53</td>
<td>1.94</td>
<td>16.8a</td>
<td>0.72</td>
<td>0.72</td>
<td>1.01</td>
</tr>
</tbody>
</table>

**Notes:** Main effects of each coping type were tested together in one model for each PTSD type. Interactions were tested in separate models, which included main effects for each risk type and each coping type. $\Delta D =$ change in deviance. *Adequate model fit. *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$
5. Results

5.1 RQ1: effects of online risk exposure on PTSD

Teens reported 222 online risk events (there were weeks when a teen would report experiencing no online risk events, and therefore did not fill out the CRIES that week). Explicit content exposure was the most common risk type (62 percent; information sharing = 15 percent, cyberbullying = 11 percent, sexual solicitation = 11 percent). While PTSD scores tended to be low (Table I), there were still teens that reported having symptoms, as the maximum score was also high across symptoms (4.75 for intrusion, 4.80 for arousal, and 5.00 for avoidance on a scale of 1 to 5). Indeed, 36.40 percent of online risk incidents resulted in clinically diagnosable PTSD (based on the cutoff score of 17 for the CRIES-13; Yule, 1998). The number of coping techniques reported on the coping behaviors checklist varied across events (active coping $M = 0.30$, $SD = 0.33$; passive coping $M = 0.55$, $SD = 0.40$; and communicative coping $M = 0.26$, $SD = 0.32$).

Fit statistics, beta weights, and significance for models testing RQ1 are provided in Table II, Part A. Please note that, based on recommendations for assessing fit for hierarchical linear modeling (Snijders and Bosker, 1994), we used change in deviance to determine if each model had adequate fit. To do this, we calculated the change in the deviance statistic when comparing the time-only model (e.g. the null model) to the final model, then used a chi-squared distribution to determine if the change in deviance was significant.

Changes in deviance from the time-only model indicated good fit, further suggesting that the frequency of online risk exposure explained variation in PTSD symptoms from week to week. Three online risk types were significantly associated with various PTSD symptoms. Explicit content exposure and cyberbullying led to significantly higher arousal symptoms, as well as marginally significantly higher avoidance symptoms. Sexual solicitation also led to significantly higher intrusion symptoms in teens. In fact, only information breaches had no significant effect on any dimension of PTSD.

5.2 RQ2: coping behaviors as a protective factor (resilience theory)

For RQ2, we did not find evidence that coping behaviors served as a protective factor against PTSD symptoms (Table II, Part B). There were significant effects, but these effects were contrary to resiliency theory and more consistent with the transactional theory of stress. Rather than reducing PTSD symptoms, main effects of coping behaviors on PTSD indicated that teens who engaged in higher than average coping behaviors experienced more PTSD symptoms, not fewer PTSD symptoms. Teens who engaged in more active coping reported more intrusion, avoidance, and arousal symptoms, while communicative coping also appeared to predict more intrusion and arousal symptoms. Thus, coping did not act as a protective mechanism to reduce PTSD, as suggested by RQ2.

5.3 RQ3: coping behaviors as a stress response

RQ3 examined whether PTSD symptoms acted as antecedents to coping behaviors (Table III). Our models indicated PTSD did predict certain coping behaviors. Active coping behaviors significantly increased as all types of PTSD symptoms increased. Teens who experienced arousal symptoms also were more likely to engage in communicative coping. Passive coping was not significantly related to PTSD symptoms (Table III), though this could be due to the low internal consistency of the measure (Kuder-Richardson $20 = 0.63$).

We also tested the interaction effects between coping behaviors and online risk exposure frequency to examine whether the association between coping and PTSD symptoms depended on the frequency of risks (Table II, Part C). These effects were also contrary to resilience theory, as coping behaviors were associated with more PTSD symptoms. Teens who engaged in a higher than average number of active coping behaviors tended to
experience more symptoms from explicit content exposure and sexual solicitation. Teens who engaged in communicative coping following sexual solicitation were more likely to report PTSD symptoms, though this relationship was stronger when risk occurrence was low. On the other hand, teens who engage in high levels of non-coping behaviors (i.e., passive coping) had lower levels of arousal and avoidance symptoms when risk factors occurred frequently. These results (see Table II, Part C) are not consistent with resiliency theory, which suggests that teens who engage in coping behaviors experience less post-traumatic symptoms as adolescent risk increases. Instead, coping behaviors were actually associated with worse outcomes, especially when risk exposure was more frequent. This also suggests that coping is used as a response to stress, rather than a protective factor, as suggested by the transactional theory of stress.

6. Discussion

Our results brought many novel insights. When examining our first research question (whether online risk exposure could lead to PTSD symptoms), we found that most online risks can lead to PTSD symptoms. Second, we found that these symptoms are not necessarily reduced by coping behaviors. Instead, as suggested by our third research question, teens tend to engage in behavior as a response to PTSD symptoms, rather than as a response to risk exposure. These results are consistent with the transactional theory of stress, which suggests that coping often occurs in response to stress, and not to the event itself. This suggests that teens tend to engage in online coping behaviors when they feel traumatized by risk exposure.

Our research was also novel in several other ways. All previous research has relied on asking teens to recall whether they had experienced an event long after it has occurred.
This method has been shown to reduce prevalence estimates in other contexts (Gorin and Stone, 2001). Thus, we found that online risk exposure was reported more often when teens were asked to recall events after a shorter period of time. The majority of teens (73 percent) who participated in the study experienced at least one risk event, which is higher than most studies of online risk exposure. In past research, between 24–57 percent of teens reported experiencing a risk event (Livingstone et al., 2010; Livingstone and Smith, 2014; Temple et al., 2012). In addition, the study took a more comprehensive approach to examining risk exposure. Previous research usually only examined one risk type in isolation (e.g. sexual solicitation or cyberbullying; Hinduja and Patchin, 2008; Temple et al., 2012) and rarely examined information sharing (Livingstone et al., 2010; Livingstone and Smith, 2014), which accounted for over a sixth of the risks teens reported.

The present study was also novel because it was the first to examine event-specific post-traumatic stress symptoms in a mainstream population (i.e. one without specific risk factors) following online risk exposure. Overall, we found that certain online risks cause more distress in teens than others, and the relationship between risk exposure and PTSD symptoms depends on what symptoms are being assessed. Cyberbullying, sexual solicitations, and explicit content exposure all had a significant effect on some type of PTSD symptoms. Information breaches (i.e. violations of privacy) had no effect on PTSD symptoms. This is consistent with previous research on teens’ online privacy; though teens may indicate that they value privacy on surveys (Lenhart et al., 2010), they tend to be less concerned with privacy than adults (Walrave and Heirman, 2011). They are more likely to share personal feelings and information online (White, 2004; Viégas, 2005). Teens are also more accustomed to having their personal contact information given to third parties without their consent (Galkin, 1996; Walrave and Heirman, 2011), and less likely to take precautions to protect their privacy (Walrave and Heirman, 2011).

Our results do suggest that other typical, weekly experiences that teens encounter online (i.e. cyberbullying, sexual solicitation, explicit content exposure) are associated with clinically diagnosable symptoms of PTSD. This is a noteworthy contribution of our work, showing the potential dark side of online engagement and social media use on adolescents. While explicit content exposure had a lesser effect on PTSD symptoms, the effect may have been weakened by our inclusion of common types of explicit content in our measures. For example, despite concern over teens’ exposure to online pornography (Kanuga and Rosenfeld, 2004), many open-ended responses in our data suggested that many teens enjoyed pornography. In addition, teens frequently exposed to explicit content may be desensitized to its traumatic effects (Cline et al., 2014), which could be another potential dark side of teen social media use. In contrast, we found no significant relationship between information breaches and PTSD, even though prior research suggests that privacy breaches make teens more vulnerable to other, more severe risks (Gross and Acquisti, 2005).

6.1 Implications for theory, policy, and design
We found that the transactional theory of stress more closely fit our data than framing coping behaviors as a protective factor as suggested by the adolescent resilience framework, as coping behaviors tended to be used in reaction to stress instead of a means to protect against it. There are several implications we can draw from these results. It is possible that in teens’ primary appraisals of an initial online risk occurrence, they did not anticipate any imminent danger and waited to engage in active and communicative coping behaviors until the risk became particularly stressful or recurred. This delayed response is consistent with risks in offline contexts. Victims often wait to take more proactive measures, such as reporting the event (Mishna and Alaggia, 2005), because they fear retaliation from the perpetrator (Camodeca and Goossens, 2005). Second, teens do exhibit coping behaviors (e.g. blocking a bully) after the potential risky and stressful situation has already occurred.
If such coping behaviors could serve as an early warning system that triggers additional resources, this may help teens navigate online risks in the long-term. For example, when a teen takes a protective action to block a contact via social media (e.g., an active coping mechanism), the site could provide context-based assistance to help teens more effectively cope with a potential risk, such as urging them to talk to a trusted adult or teaching them how to respond to bullies (Common Sense Media, 2017). It may also be possible that, while teens’ coping responses did not reduce PTSD in the short-term, they may help teens build resilience and decrease trauma symptoms over a longer period of time.

Our findings may also partially help explain why teens may not take protective measures to maintain online information privacy; privacy breaches do not prompt a strong stress response that triggers the need to cope. As such, raising teens’ awareness of the potential risks posed by oversharing may be necessary for eliciting a stress response to encourage appropriate protective measures (Madden et al., 2013). The difference in symptoms between different online risk types also has meaningful implications for legislation. While most laws focus on limiting youths’ access to explicit content (Olagunju, 2009), our results suggest that cyberbullying and sexual solicitations are more harmful to teens. Thus, it may be more beneficial to teens if legislation focused more on contact-related risks by holding individuals accountable for perpetrating these types of crimes than trying to insulate teens from content (Livingstone and Smith, 2014).

In summary, rather than focusing our efforts on trying to prevent teens from being exposed to all online risks, it may be more beneficial to teach teens more effective ways of coping with the risks they do experience (Raskauskas and Huynh, 2015). For instance, it may be helpful to teach teens how to report perpetrators to the proper authorities instead of using abstinence-based approaches that attempt to disengage teens from all social media activities. Helping teens to engage in more proactive behaviors (e.g., changing privacy filters) may not only prevent future risk exposure, it may also reduce post-traumatic stress following the event (Van der Kolk, 1994). The end goal is to allow teens to reap the positive benefits of social media engagement but avoid long-term negative effects of risk exposure.

6.2 Limitations and areas for future research

There were several limitations to our study that can inform future research. First, we based coping behaviors on a previous survey that examined teens’ behavioral responses to online risks, but later found that the passive coping measure had weak internal consistency. Future research should explore more stable techniques for measuring online coping behaviors. While measuring PTSD within a week of risk exposure was one of the strengths of our study design, we could not measure long-term PTSD symptoms that ranged beyond the duration of our study. For this reason, we encourage future research to examine longer-term effects (e.g., over a year or longer) of online risks and how teens cope with and are affected by these experiences. In addition, because many of the more severe online risks were relatively infrequent (e.g., cyberbullying), our sample size for those analyses was relatively small. We recommend that future research use a larger sample size over an extended timeframe to capture the long-term effects of online risk exposure.

The present study did not ask teens specifically what social media platform they were using when the risk occurred (though they often disclosed this in qualitative descriptions of the event). However, recent national polls of teens’ social media use and mental health outcomes suggest that certain platforms, particularly Instagram (Royal Society for Public Health, 2017), may be more detrimental than others. Certain social media sites may also have features that can facilitate more risks. For example, Facebook allows geotagging (Albrecht and McIntyre, 2015) and unmoderated video streaming (Crawford and Gillespie, 2016) that may put teens at higher risk. While our study did not address how different features of specific social media platforms impact risk exposure, this should be explored in future research.
Our sample was slightly biased toward females (63 percent). While low demographic diversity in psychological research is common (Sears, 1986), theory and research on acceptance of technology suggest that this may constrain the generalizability of our findings (e.g. the unified theory of acceptance and use of technology; Venkatesh et al., 2003). Indeed, females tend to use (Barker, 2009) and adopt (Borrero et al., 2014) social media sites for different reasons than males. Participants were also only recruited from the USA. Culture can also influence the reasons individuals adopt certain technologies (Venkatesh et al., 2003; Im et al., 2011). Thus, we recommend that future research conduct follow-up studies with larger and more diverse populations.

6.3 Conclusion
Very little research has examined the behavioral processes and psychological outcomes that occur immediately after teens experienced exposure to online risks, or the effects of or relationships between multiple types on online risk exposure and different coping behaviors in relation to post-traumatic stress. To address these limitations, we conducted a two-month diary study that found that cyberbullying, sexual solicitations, and exposure to explicit content (but not information breaches) can cause symptoms of PTSD. We also established that the transactional theory of stress was a better fit for framing the underlying processes of risk-coping (as opposed to the resilience framework). Our findings inform research related to the dark side of social media use as it pertains to teens as they experience and cope with online risks.

References
Albrecht, K. and McIntyre, L. (2015), “Psst... your location is showing!: Metadata in digital photos and posts could be revealing more than you realize”, IEEE Consumer Electronics Magazine, Vol. 4 No. 1, pp. 94-96.


Further reading


Corresponding author

Bridget Christine McHugh can be contacted at: mchugh.bridget@gmail.com

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Coping with mobile technology overload in the workplace

Pengzhen Yin
School of Management, Hefei University of Technology, Hefei, China

Carol X.J. Ou
Department of Management, Tilburg School of Economics and Management, Tilburg, The Netherlands

Robert M. Davison
Department of Information Systems, College of Business, City University of Hong Kong, Hong Kong, and

Jie Wu
School of Management, University of Science and Technology of China, Hefei, China

Abstract

Purpose – The overload effects associated with the use of mobile information and communication technologies (MICTs) in the workplace have become increasingly prevalent. The purpose of this paper is to examine the overload effects of using MICTs at work on employees’ job satisfaction, and explore the corresponding coping strategies.

Design/methodology/approach – The study is grounded on the cognitive load theory and the coping model of user adaptation. The overload antecedents and coping strategies are integrated into one model. Theoretical hypotheses are tested with survey data collected from a sample of 178 employees at work in China.

Findings – The results indicate that information overload significantly reduces job satisfaction, while the influence of interruption overload on job satisfaction is not significant. Two coping strategies (information processing timeliness and job control assistant support) can significantly improve job satisfaction. Information processing timeliness significantly moderates the relationships between two types of overload effects and job satisfaction. Job control assistant support also significantly moderates the relationship between interruption overload and job satisfaction.

Practical implications – This study suggests that information overload and interruption overload could constitute an important index to indicate employees’ overload level when using MICTs at work. The two coping strategies provide managers with effective ways to improve employees’ job satisfaction. By taking advantage of the moderation effects of coping strategies, managers could lower employees’ evaluation of overload to an appropriate level.

Originality/value – This study provides a comprehensive model to examine how the overload resulting from using MICTs in the workplace affects employees’ work status, and how to cope with it. Two types of overload are conceptualized and corresponding coping strategies are identified. The measurements of principal constructs are developed and empirically validated. The results provide theoretical and practical insights on human resource management and human–computer interaction.

Keywords Information overload, Job satisfaction, Coping model of user adaptation, Coping theory, Interruption overload, Mobile information and communication technologies (MICTs)

Paper type Research paper

1. Introduction

Technology overload, including mobile application overload, information overload, communication and interruption overload, has become increasingly prevalent in the digital workplace. A recent survey conducted by RingCentral found that: more than 70 percent of employees say their communication volume is a challenge to the fulfillment of their work; 68 percent of employees toggle between mobile applications up to ten times per hour; 31 percent of employees indicate toggling causes them to lose their train of thought; and 56 percent of workers find searching for information from multiple sources disruptive.
Workplace information and communication technologies are meant to help employees work more effectively and keep team members more focused in a collaborative way so that they can be more productive. However, employees are becoming more frustrated due to the extent to which they are overloaded by technology (Brumberg, 2018).

In particular, mobile information and communication technologies (MICTs) (e.g. smartphones, tablets and mobile application software) have been considered to help users overcome the physical strains associated with traditional business interactions (Cao et al., 2016; Jeske and Axtell, 2014). On the one hand, the permanent connectivity of MICTs can increase the extent to which employees need to process multiple information demands (Jarvenpaa and Lang, 2005). On the other hand, the continuous connectivity of MICTs might induce individuals' feeling of cognitive overload, which is particularly significant in an era of information explosion (Hung et al., 2011).

Since people make extensive use of MICTs, excessive engagement in interactions with MICTs has become commonplace. Considering the ubiquity of MICTs, it is reasonable to infer that the problematic mobile technology overload is one contributing factor to a reduction in employees' productivity and job satisfaction. Therefore, it is necessary to investigate the sources and the extent of technology overload caused by the use of MICTs in workplace, and examine its effects on employees' job satisfaction. Accordingly, we propose the following two research questions:

RQ1. Which technology overload factors exist and how will they affect employees' job satisfaction?

RQ2. Which coping strategies will decrease the overload effects of using mobile technologies in the workplace, and how?

In this study, through an extensive review of the literature, we identify two types of technology overload: information overload and interruption overload. Furthermore, based on coping theory and the characteristics of mobile technologies, we propose two specific mechanisms that may reduce individual perceived negative effects of MICTs usage in the workplace: information processing timeliness and job control assistant support. The coping model of user adaptation (CMUA) (Beaudry and Pinsonneault, 2005) is used as a framework to establish the research model.

The remainder of the paper is organized as follows. In Section 2, we review the literature about the technology overload phenomenon and associated coping strategies. Based on the literature review and theoretical foundations, we build our conceptual model and develop corresponding hypotheses in Section 3. In Section 4, we explain the survey research methods of this study followed by the data analysis. Then, we discuss the findings and suggest future research before concluding the paper with implications and contributions.

2. Literature review

2.1 Technology overload

In the era of the knowledge economy, the prevalently cooperative tasks undertaken by employees demand more effective and efficient information processing and communication in the organization (Mäntymäki and Riemer, 2016). Increasingly information systems, devices and applications for organizational or individual use are developed to support the requirements of human communication. The multiple sources of information and our communication requirements result in an increasing level of perceived overload by individuals. This phenomenon has been described as “technology overload” in the literature, and has been defined as “device proliferation and/or information overload that causes cognitive and physical burdens on human beings due to the use of multiple gadgets with multiple functions to accomplish multiple tasks in everyday activities” (Grandhi et al., 2005).
Technology overload has been found to reduce individual productivity (Karr-Wisniewski and Lu, 2010), and increase the feeling of stress and decrease job satisfaction (Ragu-Nathan et al., 2008; Tarafdar et al., 2011). Typically, in the technostress literature, this phenomenon has been operationalized as one construct, i.e. techno-overload (Ragu-Nathan et al., 2008). Hung et al. (2011) extended this stream of research to the ubiquitous technostress context and investigated the technology overload effects on job stress and individual productivity.

Scholars have recently investigated the technology overload phenomenon from two perspectives. The first is in the context of general organization IT environment, which focuses on organization supported technology (Karr-Wisniewski and Lu, 2010). The second examines the increasingly prevalent phenomenon of IT consumerization and Bring Your Own Device in the workplace (Yun et al., 2012).

In the first stream of research, Karr-Wisniewski and Lu (2010) proposed three components of technology overload: information overload, interruption overload and system feature overload. The system feature overload “occurs when the addition of new features is out-weighed by the impact on technical resources and the complexity of use.” That is, the system feature overload only happens in situations where the provided IT features, interfaces or functions exceed users’ ability to cope with them or they are too complex to use.

In this study, we focus on the context of employees’ proactive adoption of mobile technologies in workplace, where employees are “more aware of technology in the workplace and able to choose software and devices that are optimally suited to their work” (Niehaves et al., 2013). Further, employees usually consider their own consumer devices and apps as easier and more intuitive to use (Harris et al., 2012). Therefore, employees are more familiar with these kinds of devices due to work-life dual usage behavior (Niehaves et al., 2013). The mobile technologies used for work purposes have been voluntarily and spontaneously selected by employees based on their own work practices and personal habits. The exploitation of privately gained competences enhanced their capability to cope with these mobile technologies at work (Niehaves et al., 2013). As a result, the above-mentioned system feature overload is not considered an overload issue any more.

In addition, to the best of our knowledge, only one paper refers to the concept of “system redundancies” (Ortbach et al., 2013). In their qualitative research work, four keywords are used to reflect the existence of system redundancies, which are “frequent changes of system, increased multi system usage, lack of comfort, redundancies of data” (Ortbach et al., 2013). They claim that the main stressor in the system redundancies dimension is the “necessity to frequently switch contexts between multiple different channels.” Therefore, this view is consistent with the research of Zhang and Rau (2016) who considered that the technology per se is the primary source of multitasking. Further, Aral et al. (2012) suggested that multitasking creates high cognitive switching costs. Therefore, the negative effect of system redundancy constitutes a cognitive overload due to the large amount of information that needs to be processed and the repeated interruption requests due to the frequent switching among multiple tasks through multiple technologies.

Thus, in this study, we focused on two primary negative effects of MICTs: information overload originating from multiple communication channels and interruption overload caused by requests enabled by MICTs.

Literally, the term “information overload” refers to a person receiving too much information (Eppler and Mengis, 2004). Researchers initially focused on investigating the inverted U-shaped relationship between individual performance and the received amount of information. When the received amount of information exceeds a certain point, individual performance will decline (Chewning and Harrell, 1990).

The definition of information overload also refers to a comparison between an individual’s information processing capability and information processing requirements.
That is, the information overload will occur when the information processing requirements are greater than the information processing capabilities. The “capabilities” and “requirements” can be measured in terms of a certain time period. In addition, researchers have also claimed that the available processing time is an important element that causes information overload (Schick et al., 1990; Owen, 1992; Iselin, 2010). In the information systems literature, information overload is treated as individuals’ subjective experience. Scholars have indirectly investigated the level of information overload by empirically examining individuals’ perceived levels of stress, anxiety or pressure (Haksever and Fisher, 1996).

Thus, in the current literature, the phenomenon of information overload involves two issues. First, it highlights the fact that individuals often receive a large amount of information. Second, it recognizes that individuals will experience information overload when their information processing capability is insufficient given the amount of information received. However, the adoption and widespread use of mobile technologies in the workplace means that the information overload phenomenon has become increasingly prevalent. The ubiquitous nature of mobile technologies provides users with multiple sources of information communication and the ability to be continuously connected. These functions cause users to be exposed to massive amounts of information, which enhances the perception of information overload.

In the current literature, there is no consistent definition of the concept of interruption overload. The term “interruption” refers to “any distraction that makes an individual stop his/her planned activity to respond to the interrupt’s initiator” (Jackson et al., 2001). The interruption overload construct is different from information overload because it highlights the frequent attention distractions in a given circumstance. In practice, individuals are required to process information from both internal and external sources in parallel, which results in high levels of interruption (Ragu-Nathan et al., 2008). Therefore, in this study, we defined the interruption overload as the degree to which MICT users experience an overload of disturbance from unscheduled MICT interactions, or the discontinuity of current work activity because of MICT interactions that are not initiated by the focal employee.

It is worth noting that communication overload and interruption overload overlap in the literature. Frequently occurring communications are a major cause of interruptions (Ou and Davison, 2011). Thus, interruption overload occurs together with communication overload. Although these overloads may affect individuals in different ways, the two concepts are highly correlated. However, few studies have been conducted to distinguish the antecedents and impacts of these phenomena. In this study, we focus on the influence of interruption overload on individual job satisfaction.

The use of MICTs in the workplace causes interruption overload to increase; this is thus a topic that is worthy of more research attention. In the literature, researchers have investigated the influence of the synchronicity of IT on individual communication behavior. MICTs have been identified as highly synchronous technologies (Carlson and George, 2004). The existence of multiple communication channels supported by MICTs means that users may be frequently interrupted by unscheduled interactions, which require more cognitive resources to change among diverse mental models, and eventually result in high levels of perceived overload. On the other hand, the high synchronicity of MICTs means that users do not have enough time to effectively integrate information into a specific mental model. In this way, the frequent discontinuity of current tasks will also increase an individual’s perceived interruption overload (Garrett and Danziger, 2008; Ou and Davison, 2011).

2.2 Coping strategy
Coping has been theorized as an effective way to reduce an individual’s perceived stress level (Lazarus, 1993). From the perspective of coping theory, some researchers have also
investigated coping strategies when individuals experience overload that results in stress or other negative consequences (Lazarus, 1993; Weinert et al., 2013). Coping theory incorporates two components: problem-focused coping and emotional-focused coping.

Specifically, problem-focused coping refers to redefining problems, generating alternative solutions, weighting the alternatives in terms of individuals' costs and benefits, and choosing the most appropriate alternatives so as to be able to act (Lazarus and Folkman, 1984). Research has shown that the person who uses a problem-focused coping strategy may more easily avoid the strains that occur between work and life (Lapierre and Allen, 2006). Emotional-focused coping consists of actions or thoughts to control the undesirable feelings that result from stressful circumstances (Thoits, 1986). However, while emotional-focused coping can change individuals' perception of the environment, it does not change the external environment itself (Beaudry and Pinsonneault, 2005).

According to the literature in the field of mobile technology usage, MICTs provide multiple functions to facilitate users' demands and requirements in a specific context by personalizing these MICTs. For example, users can add context-based applications and software, or configure personal tools for work. Some mobile applications have also been developed to help users manage their schedule, such as notes, clock, timer, caller ID, reminders and so on (Sarwar and Soomro, 2013). In other words, MICTs enable users to actively adopt available applications to organize their lives and therefore may eventually release their sense of being strained. According to coping theory, these supporting functions of MICTs are problem-focused coping strategies for users that have the potential to enhance individual work effectiveness.

In addition, MICTs can also help users cope with the negative effects by providing timely information. Previous studies in the mobile technology design and use fields have widely referred to the timeliness characteristics of MICTs. Basole and Chao (2004) defined timeliness as the degree to which a system provides a user with current and appropriate information; they concluded that timeliness is an important contributor to the perceived usefulness of mobile technologies. That is, the timeliness characteristic of MICTs fits users' demands for efficient information provision and processing. Based on this stream of literature, MICTs provide users with solutions to cope with potential strains in situations characterized by tight time limits or parallel task requirements. Therefore, we propose that the timeliness characteristic of MICTs could constitute an effective problem-focused coping strategy for users given the information timeliness and communication efficiency enhanced by using MICTs (Liang et al., 2007).

### 2.3 Coping model of user adaptation

The CMUA was first proposed by Beaudry and Pinsonneault (2005). This theoretical framework is grounded on coping theory (Lazarus, 1993). The user adaptation concept is illustrated in light of the coping theory in the CMUA framework, and defined as the “cognitive and behavioral efforts exerted by users to manage specific consequences associated with a significant IT event that occurs in their work environment” (Beaudry and Pinsonneault, 2005). Two appraisal processes have been included in this model: primary and secondary appraisal. The primary appraisal is triggered by a significant IT event that an individual evaluates as an opportunity or a threat. In the second appraisal, individuals first evaluate their personal control of the IT event and second consider the adaptation options for them. Therefore, according to the two appraisal processes, four kinds of adaptation strategies are proposed, namely, benefits maximizing, benefits satisficing, disturbance handling and self-preservation. After the adaptation, there are different outcomes. The comprehensive framework of the CMUA (Beaudry and Pinsonneault, 2005) is shown in Figure 1.

The CMUA framework demonstrates how individuals can take different outcome routes after the two appraisal processes. Few studies have explored the negative
consequences of IT, and even fewer researchers have investigated the positive outcomes of user adaptation to the negative consequences of IT use. According to the CMUA framework, while some users evaluate an IT event as a threat, those who have a high degree of control over the IT event adopt disturbance handling adaptation strategies, which may result in the improvement of individual efficiency and effectiveness. This route shows that a perceived threat turns out to enhance positive outcomes of employees by conducting appropriate adaptation strategies or coping strategies, which have been rarely studied in the literature.

In addition, according to CMUA, the outcome factors of user adaptation to the IT event are also worth exploring. IT applications exert negative impacts not only on human physical health but also on employees’ psychological states (Lee et al., 2014). The form and nature of work have been significantly changed as a result of MICT usage. Technology overload effects associated with the use of MICTs have also become increasingly significant. The new work forms may cause employees to experience a higher density of information exchange and greater negative impacts from work burdens. However, few studies have examined the effects of technology overload on aspects of employees’ psychological status at work.

Current studies have treated job satisfaction as one of the main objectives of an organization that is undertaking the development and adoption of IT (Weiss and Leimeister, 2012). Locke (1976) first defined job satisfaction as “a pleasurable or positive emotional state resulting from the appraisal of one’s job or job experiences.” Therefore, job satisfaction has been treated as an important psychological output variable, and has also been widely studied in the technostress literature because of its influence on both employee and organization performance (Suh and Lee, 2017). Thus, job satisfaction is selected as the main outcome factor in this study.

To summarize, according to the above literature review, we adapted the CMUA framework to build our conceptual model of technology overload. In the context of MICT usage in the workplace, we propose two important technology overload factors and two problem-focused coping strategies that reflect individual efficiency and effectiveness. In addition, we also examine the moderation effects of two adaptation strategies on the relationship of perceived technology overload and psychological outcomes.

**Figure 1.** Coping model of user adaptation (CMUA) framework

Source: Beaudry and Pinsonneault (2005)
3. Hypothesis development

In this section, grounded on the CMUA framework and coping theory, we explain our research model and hypothesis development in detail. The proposed research model is shown in Figure 1. In the research model, we explore the relationships among the technology overload factors (i.e. information overload and interruption overload), coping strategies (i.e. information processing timeliness and job control assistant support) and job satisfaction. Table I lists the definition of each construct used in this study.

3.1 The effects of information overload

Information overload is an increasingly significant problem encountered by individual employees. The continuous experience of a high level of information overload resulting in high levels of cognitive load is one of the main sources of job stress (Folkman et al., 1986). The use of MICTs in the workplace has made the working environment more complex and information rich (Bawden and Robinson, 2009). Individuals need to quickly respond to information processing requirements and have to identify useful information to best fit the task requirements and make related decisions or conclusions based on this information. Employees are often requested to effectively extract information that needs continuous concentration and more cognitive resources. This environment causes individuals to experience different degrees of negative effects, which directly reduce the positive emotions associated with work. Therefore, information overload is an important factor that can reduce individuals’ job satisfaction. We thus hypothesize that:

\[ H1. \text{ Information overload has a negative impact on job satisfaction.} \]

3.2 The influence of interruption overload

A few studies have discussed the negative effects of interruption resulting from IT use in the workplace (Garrett and Danziger, 2008). Specially, overload effects of interruptions may result in negative impacts on an individual’s life, such as increasing stress levels and leading to inefficiency (Karr-Wisniewski and Lu, 2010). In fact, knowledge workers experience a high frequency of interruptions, since with the proliferation of MICTs, employees require almost 5–15 additional minutes to regroup for productive thinking (Rigby, 2006). The high level of interruption overload results in less productive employees. Frequently switching between multiple tasks also leads to an increasing level of negative feelings, such as stress.

<table>
<thead>
<tr>
<th>Principal variables</th>
<th>Definitions</th>
<th>Sources and scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information overload (IO)</td>
<td>The degree to which MICT users feel that the amount of information exceeds their limited human information processing capacity</td>
<td>Adapted from Eppler and Mengis (2004)</td>
</tr>
<tr>
<td>Interruption overload (ITO)</td>
<td>The degree to which MICT users experience an overload of disturbance from unscheduled MICT interactions, or the discontinuity of current work activity because of MICT interactions which are not initiated by the focal person</td>
<td>Adapted from Ou and Davison (2011)</td>
</tr>
<tr>
<td>Job satisfaction (JOBSA)</td>
<td>A pleasurable or positive emotional state resulting from the appraisal of one’s job or job experience</td>
<td>Adapted from Locke (1976)</td>
</tr>
<tr>
<td>Information processing timeliness (IPT)</td>
<td>The degree to which an individual perceives the support from MICTs for timely information processing</td>
<td>Developed in this study based on Liang et al. (2007)</td>
</tr>
<tr>
<td>Job control assistant support (JCAS)</td>
<td>The degree to which an individual perceives the support from MICTs by using related functions to help himself/herself influence or manipulate the work process</td>
<td>Developed in this study based on Hung et al. (2011)</td>
</tr>
</tbody>
</table>

Table I. Principal variables and definitions
or frustration (Eyrolle and Celler, 2000). Thus, interruption overload can directly cause
individuals to experience negative emotional feelings. Accordingly, we hypothesize that:

H2. Interruption overload has a negative impact on job satisfaction.

3.3 Coping strategies

3.3.1 The influence of information processing timeliness. Timeliness has been identified as
an important indicator to measure the fit between task and technology in the literature
(Lee et al., 2007). Timeliness has also become an important indicator of the effectiveness and
usefulness of information systems (Sharda et al., 1988). With the development of the global
economy, traditional forms of work have been changed. Individuals have to communicate
and coordinate with others who are geographically remote and may be located across
different time zones. These increasingly complex work environments require individuals to
acquire and process information in a timely fashion in order to support decision making.
Furthermore, in the mobile context, the importance of timeliness has been amplified
(Basole and Chao, 2004). Information timeliness and communication efficiency have become
two of the important affordances of mobile technologies, and have improved individual
productivity and organizational profitability (Liang et al., 2007). These affordances of mobile
technology are attributed to the synchronous characteristic of MICTs (Ragu-Nathan et al.,
2008), which enable employees to simultaneously deal with multiple streams of information
that may come from different sources and be related to different tasks (Straus et al., 2010).
Thus, the information timeliness provided by mobile technologies is also closely related
to end-user satisfaction (Zviran et al., 2006). ICTs support users by providing the ability to
process information quickly and more effectively, and improving the perceived quality of
their jobs. Therefore, as a high synchronicity technology, MICTs effectively support users
with information processing timeliness. Thus, we hypothesize that:

H3. Information process timeliness improves an individual’s job satisfaction.

3.3.2 The effects of job control assistant support. Based on the task technology fit theory
(Goodhue and Thompson, 1995), the fit among task requirements, technological
characteristics and personal capabilities can lead to enhanced performance (Liang et al.,
2007). Therefore, many intrinsic functions of MICTs have been designed to facilitate the
interaction between humans and MICTs, such as filters for spam phone calls, job-task
management, calendar reminders, etc. In this study, we regard these supporting functions as
forms of job control assistant support in the work context. These supporting functions of
MICTs help users to gain better control of their daily communication and task arrangements.
Furthermore, these functions can also provide users with positive psychological hints for the
high level of perceived control. That is, on the one hand, the functions of job control assistant
support provide individuals with a more effective form of work. On the other hand, the these
functions can positively affect users’ psychological states due to the enhancement of the
feeling of control, and the uncertainty reduction in the workplace, which directly improve an
individual’s job satisfaction (Ragu-Nathan et al., 2008). Therefore, we hypothesize that:

H4. The functions of job control assistant support are positively related to an
individual’s job satisfaction.

3.4 Moderating effects of coping strategies
Researchers have also identified various kinds of moderators, such as personality features,
social environment and organizational mechanisms (e.g. social support) in the field of
technostress (Fuglseth and Sorebo, 2014). However, the results of the interaction effects are
diversified and inconsistent (Ragu-Nathan et al., 2008). Further, aligned with the coping theory, plenty of research has been conducted to investigate the moderating effects of coping strategies on the stress–outcome relationships (Tidd and Friedman, 2002). However, few studies have been conducted to examine the specific moderating effects of coping strategies on the technology overload–outcome relationship. Therefore, in this section, we aim to investigate the moderating effects of the two proposed problem-focused coping strategies in the MICT context on the relationship between technology overload factors and job satisfaction.

3.4.1 The moderating role of information processing timeliness. Information processing timeliness has been identified as a factor that can improve an individual’s work efficiency when using information technologies in the workplace (Gattiker and Goodhue, 2004). Typically, increasing information overload has led to an information processing environment that is very demanding (Tushman and Nadler, 1978). If the information is processed effectively and efficiently, the feeling of information overload can be reduced. On the one hand, MICTs have facilitated employees’ work by changing the spatial distance, which enables employees to respond to communication requests without any geographical constraint in a timely fashion. On the other hand, MICTs also enable users to transform the information in a timely manner by providing multiple communication channels, thus bridging the temporal distance between interlocutors. Therefore, the timely manner of information processing can enhance employees’ feeling of control of multiple tasks and thus reduce the feeling of information overload. Therefore, we hypothesize:

\[ H5a. \text{ Information processing timeliness negatively moderates the relationship between information overload and job satisfaction, meaning information processing timeliness can reduce the negative effect of information overload on job satisfaction.} \]

The contemporary work environment has changed a lot with the development and adoption of IT. Work tasks have become increasingly complex with the uncertainty of the market. These complex work tasks involve high cognitive load and may be easily susceptible to interference from frequent interruptions (Speier et al., 2003). Specifically, the use of MICTs in the workplace may increase the complexity of work tasks, because various kinds of information can be transmitted through multiple channels, and individuals are busy handling overloaded information from multiple sources. It has been recognized that the use of MICTs in the workplace has resulted in increasing levels of interruption, such as cell phone interruptions (Avrahami et al., 2007). In this context, the information processing timeliness function supported by MICTs facilitates users to complete their tasks more quickly. Therefore, when an interruption occurs, the fast information processing requirements and information processing timeliness facilitation will increase individuals’ perceived cognitive load level, because complex tasks challenge individuals’ cognitive capacity. In these cases, MICT users experience high levels of cognitive strain and feel exhausted. Therefore, according to cognitive dissonance theory (Bhattacherjee and Premkumar, 2004), individuals may change their beliefs or attitudes to their current jobs, resulting in an even higher level of negative feelings. Therefore, we hypothesize that:

\[ H5b. \text{ Information processing timeliness positively moderates the relationship between interruption overload and job satisfaction, meaning information processing timeliness can further enhance the negative effect of interruption overload on job satisfaction.} \]

3.4.2 The moderating role of job control assistant support. The functions of job control assistant support enabled by MICTs are designed to provide users with more control over the information that they receive and send. These intrinsic functions of MICTs have improved individuals’ work effectiveness in the sense that MICTs help users work in an organized way. In the context of information overload, it is inevitable that MICT users must
process a larger amount of information. The function of job control assistant support can help users identify the extent to which the information is important or list the tasks according to their priority. Thus, although the absolute amount of information may increase for the employee, the technology-enabled ability for employees to arrange the work list and manage work-related uncertainty can nurture the feeling of control at work. These helpful functions may result in positive feelings about their job. Thus, we hypothesize that:

**H6a.** Job control assistant support negatively moderates the relationship between information overload and job satisfaction, meaning job control assistant support at MICTs can reduce the negative effect of information overload on job satisfaction.

On the other hand, the control enabled by the job control assistant support of MICTs may facilitate users to effectively manage their personal tasks and other daily affairs. That is, users would be informed in advance about potential interruptions and so would be able to prepare for corresponding coping strategies. Therefore, the function of job control assistant support of MICTs may significantly reduce individuals’ strain level and we hypothesize that:

**H6b.** Job control assistant support negatively moderates the relationship between interruption overload and job satisfaction, meaning job control assistant support at MICTs can reduce the negative effect of interruption overload on job satisfaction.

In addition, to investigate the influence of the focal constructs in the proposed research model, we also control for the effect of differences in five individual characteristics on job satisfaction in the context of mobile technology use at workplace: age, gender, education, position and work experience. We select these five controls following the literature in the field of technostress. Specifically, demographics including gender, age, education and work experience can lead to different degrees of work-life conflict (Ahuja et al., 2007), which may directly influence individual job satisfaction. It appears reasonable that employees with different individual characteristics may experience different levels of stress in the workplace, which may indirectly influence individual job satisfaction (Ragu-Nathan et al., 2008). Consistently, based on our initial interview results, the patterns of using mobile technologies at workplace differ across different positions and age groups. Thus, we include these individual demographics in the research model so as to control their effects on job satisfaction. The proposed research model is then summarized in Figure 2.

**Figure 2.** Research model
4. Methodology

In this study, we went through several procedures to empirically develop and validate the instrument for technology overload factors and coping strategy factors. First, following the literature review, we developed the measurements for the constructs, which we pre-tested through a series of procedures including card sorting and ranking (Moore and Benbasat, 1991). We then conducted a survey to validate the instrument with 178 valid data points. We detail those procedures below.

4.1 Instrument development and questionnaire design

The literature does not provide consistent definitions about the constructs of information overload, information processing timeliness and job control assistant support. Therefore, we clarified the definitions and the developed measures of these three constructs based on the literature review. The corresponding definitions and measurements for other two constructs, namely, interruption overload and job satisfaction, are adapted from prior studies.

To ensure the face and content validity of the definitions, we showed the definitions to ten randomly selected practitioners and invited them to comment on the fitness of these definitions with their work practice. According to their feedback, we revised the definitions. Then, based on these definitions and literature, we developed a list of initial items in a pool to measure the constructs of information overload, information processing timeliness and job control assistant support. The interruption overload was operationalized based on the conceptual definition which has been proposed in Ou and Davison (2011). The measures of job satisfaction are adapted from Ragu-Nathan et al. (2008) and Spector (1985). The initial constructs validity and reliability for the instrument are established according to the procedures introduced by Straub (1989).

Following the processes used in prior studies (Moore and Benbasat, 1991), two rounds of card sorting were conducted to further test the constructs’ reliability and validity. First, four judges (PhD students) were presented with the items without the labels and randomly sequenced. They were asked to perform two tasks: group the related items into several categories; and label and define each category that they have classified. The correct hit ratio was 79 percent. The results provide initial evidence for construct validity. The items that are classified into the same group indicated convergent validity. The distinct dimensions that emerged from this process demonstrate the discriminant validity. The items with multiple classifications were revised and used for the next analysis. We then revised some wordings that were ambiguous and conducted a second round of card sorting. Four new judges, including two research students and two IT professionals (one is an IT maintenance management employee and the other one is a CIO’s assistant), were enrolled. The construct names were provided in this round and the correct hit ratio was 94 percent, indicating sufficient item-construct reliability (Moore and Benbasat, 1991). All the selected judges fit our research context in the sense that they are MICT users, and they frequently use MICTs to deal with their work-related tasks.

We notice that only two items were included in the measure of information processing timeliness after the card sorting exercise. In order to examine the validity of using a small number of items, we looked for support from past studies. We found that Verhoef (2013) used two items to measure “payment equity.” Consistently, Eisinga et al. (2013) comprehensively analyzed the reliability of a two-item scale, and pointed out that the most appropriate reliability coefficient for the two-item scale is the value of Spearman–Brown statistic and coefficient α, rather than Pearson correlation. Therefore, we calculate the Spearman–Brown value and coefficient α of “information processing timeliness” as suggested in their research. The results shown that the Spearman statistic value is 0.535 ($p < 0.01$) and the coefficient α equals to 0.66, which are both acceptable. Furthermore, Bergkvist and Rossiter (2007) have demonstrated that there is no difference in predictive validity between single-item and
multiple-item measures, especially for singular and concrete constructs. The concept of “information processing timeliness” is essentially a singular construct, because it reflects the effects of time factors during information processing. Thus, the two-item construction of this construct should not reduce its predictive validity. As a result of the above scrutiny, we decide to take the two-item scale to measure “information processing timeliness” in this study, but we meanwhile acknowledge in the future research that this scale can be further improved.

After the above procedures, a pilot study was conducted with a sample of 30 Chinese graduate students from a university in Mainland China. We further revised the questionnaire based on the results and feedback of the pilot test. The final questionnaire (as shown in Table A1) was used for large-scale data collection.

4.2 Data collection

This research aims to study the employees who are using MICTs for work purposes. To ensure that we could include respondents from a variety of backgrounds, we collected the data from two sources simultaneously. First, we conducted a web-based survey. The questionnaire link was randomly sent to several qualified respondents located in different cities in China (i.e. Beijing, Shenzhen, Nanjing and Chengdu), and then the web-based survey was disseminated using a snowball method to the colleagues of respondents in the first round in this web-based survey. In total, 98 valid data points were collected. Second, we sent the paper-based questionnaire to working professionals who are also undertaking part-time post-graduate study at a university in Hefei, China. With the paper-based survey, we collected 80 valid respondents, with a response rate of 89 percent. All of the respondents have adopted MICTs in their workplace and most of them are working in the internet industry.

We verified that the non-response bias was not a concern using the methods described in Armstrong and Overton (1977). The demographic characteristics of the respondents from both of these sources were similar. A t-test of the demographic characteristics of respondents who submitted their answers in the first two weeks and in the last two weeks did not significantly differ (p > 0.05). It is worth noting that sample overlapping did not appear to be a significant issue in this study. When the web survey is answered, the hosted website (a third-party platform) automatically captures the IP address and a duplicate of this IP address is not allowed for a second response. Each student in the MBA class can answer only one paper-based questionnaire. Furthermore, we collected data in different cities at almost the same time. It is very unlikely that one person could answer both the web-based and the paper-based questionnaire. Hence, the 178 data points were used in the subsequent statistical analysis. The demographic characteristics of the sample are shown in Table II.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Items</th>
<th>%</th>
<th>Categories</th>
<th>Items</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>16–20</td>
<td>1.7</td>
<td>Gender</td>
<td>Male</td>
<td>53.4</td>
</tr>
<tr>
<td></td>
<td>21–25</td>
<td>24.2</td>
<td></td>
<td>Female</td>
<td>46.6</td>
</tr>
<tr>
<td></td>
<td>26–30</td>
<td>55.6</td>
<td>Working experience (years)</td>
<td>&lt;1</td>
<td>30.9</td>
</tr>
<tr>
<td></td>
<td>31–35</td>
<td>12.9</td>
<td></td>
<td>1–3</td>
<td>31.5</td>
</tr>
<tr>
<td></td>
<td>36–40</td>
<td>5.1</td>
<td></td>
<td>3–5</td>
<td>15.7</td>
</tr>
<tr>
<td></td>
<td>41–45</td>
<td>0.6</td>
<td></td>
<td>5–7</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td>&gt; 45</td>
<td>0</td>
<td></td>
<td>7–10</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt; 10</td>
<td>6.2</td>
</tr>
<tr>
<td>Education</td>
<td>High school</td>
<td>1.7</td>
<td>Position</td>
<td>Non-manager</td>
<td>61.8</td>
</tr>
<tr>
<td></td>
<td>Associate degree</td>
<td>3.9</td>
<td></td>
<td>General manager</td>
<td>23.6</td>
</tr>
<tr>
<td></td>
<td>Undergraduate</td>
<td>55.6</td>
<td></td>
<td>Middle manager</td>
<td>12.4</td>
</tr>
<tr>
<td></td>
<td>Graduate</td>
<td>38.8</td>
<td></td>
<td>Top manager</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Table II. Demographic characteristics

Note: n = 178
Our respondents all work in the internet industry. As shown in Table II, most of the respondents are in the age group between 21 and 35, and only 5.7 percent are aged from 36 to 45. Such an age distribution is consistent with the current information about employees in the internet industry (see Table AII). The work experiences of the employees are reasonably diverse. Our sample also includes all the subgroups of education level including high school, associate degree, undergraduate and graduate, with the majority having undergraduate and graduate qualifications. The ratio of females and males in the sample is almost equal. The ratio of managers and non-manager employees is also approximately equal. Thus, although our sample is relatively young, these respondents are distributed in different work positions and also cover different degrees of work experience. Thus, our respondents can be considered reasonably representative of the sample in the internet industry.

5. Data analysis and results

5.1 Validating the measures

SPSS 22.0 was employed to calculate the construct validity and reliability. First, we conducted the exploratory factor analysis (Fabrigar et al., 1999). As shown in Table III, all factor loading scores on their expected factors are above 0.7. The factor loading scores on their expected factors are higher than those on other factors, which indicates healthy discriminant validity. All the communality scores are higher than 0.5. Based on these results, we have demonstrated the adequate validity of the measures.

Second, we conducted a confirmatory factor analysis using SmartPLS 2.0. The corresponding results are shown in Tables IV and V. The factor loadings of all items with CFA approaches are larger than 0.7. The construct reliabilities for all principal constructs are assessed by identifying the composite reliability scores, all of which are above 0.8 and so are greater than the recommended minimum value of 0.7. The average variance extracted (AVE) values ranged from 0.6934 to 0.7743 and are larger than the suggested value of 0.5. These results demonstrate the acceptance of convergent validity. In addition, all Cronbach’s α...
values are greater than the recommended minimum value of 0.6, which demonstrates the reliability of the measures (Lyberg et al., 1997). As shown in Table V, the square roots of AVE are all above 0.8, and are greater than all other cross-correlations. This result indicates that each construct captures more expected construct variance than error variance. That is, the discriminant validity of the measures is good. In all, the above-mentioned results demonstrate adequate reliability, and convergent and discriminant validity for all constructs used in this study.

Furthermore, we have subjected the data to tests of common methods bias. As shown in Table III, the exploratory factor analysis results have revealed a five-factor structure (all the eigenvalues of the five factors are greater than 1), and all the factors roughly explain equal variance (9.633–20.508 percent) in the data, reflecting the lack of substantial common method variance (Anand et al., 2010). In addition, the correlation matrix in Table V indicates that the highest inter-construct correlations are below 0.56, thereby also reflecting the lack of substantial common methods bias, since common methods bias is usually exhibited by extremely high correlation among constructs ($r > 0.9$) (Bagozzi et al., 1991). Finally, we conducted the multicollinearity diagnostics for these constructs using SPSS. The commonly accepted standard to evaluate the existence of multicollinearity is that the tolerance value is less than 0.1, or the variance inflation factors (VIFs) are greater than 10 (Kutner et al., 2004). The results show that the lowest tolerance value was 0.408, and the highest VIF was 2.45. Thus, the multicollinearity did not seem to be a significant problem in our data set.

### 5.2 Testing the conceptual model

To test the proposed research model, statistical procedures described by Baron and Kenny (1986) were conducted. We first created the multiplicative interaction terms between technology overload factors and each of the coping strategy factors. Second, hierarchical regression model analyses were carried out, including multiplicative interaction terms. All the control variables (age, gender, education, position and work-age) were put into the model and named as Model 1. In Model 2, both the control variables and technology overload factors are selected into the model. The problem-focused coping variables are added into the Model 2, and labeled as Model 3. Both Models 2 and 3 indicated the main effects of
independent variables on job satisfaction. In Model 4, the multiplicative interaction terms (coping strategy factors × technology overload factors) were entered. The significance of interaction terms indicates the existence of moderating effects; otherwise, there will no moderating effects (Osborne and Costello, 2004). The results of the hierarchical regression analysis are shown in Table VI.

In Table VI, Model 4 shows the proposed research model results. According to Model 4, we can see that the information overload is significantly and negatively related to job satisfaction ($b = -0.147$, $p < 0.05$), supporting $H1$. There are no significant effects between interruption overload and job satisfaction ($b = 0.023$, $p > 0.05$), rejecting $H2$. Both the coping strategy factors significantly improve individuals’ perceived job satisfaction level (for information processing timeliness $b = 0.217$, $p < 0.01$; and for job control assistant support $b = 0.515$, $p < 0.001$, respectively). Therefore, both $H3$ and $H4$ are supported. The significances of the interaction items show the results of the moderating effects. The negative moderation effect of information processing timeliness on the relationship between information overload and job satisfaction is significant ($b = -0.151$, $p < 0.05$), supporting $H5a$. The positive moderation effect of information processing timeliness on the relationship between interruption overload on job satisfaction is also significant ($b = 0.366$, $p < 0.001$), supporting $H5b$. On the other hand, the moderation effect of job control assistant support on the relationships between information overload and job satisfaction is supported by our data, though not significant ($b = 0.131$, $p > 0.05$), thus rejecting $H6a$ and $H6b$. Job control assistant support demonstrates a negatively moderating effect on the relationship between interruption overload and job satisfaction ($b = -0.269$, $p < 0.001$), supporting $H6b$. It is worth to note that by adding the information processing timeliness and job control assistant support into the Model 3 and then their moderating effects in Model 4, both values of $R^2$ show a significant increase. This result indicates that the problem-focused coping functions supported by the MICTs significantly changed individuals’ attitude to their job by improving their work efficiency and effectiveness.

<table>
<thead>
<tr>
<th>Model</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DV = job satisfaction</strong></td>
<td><strong>Control variables</strong></td>
<td><strong>Main effects</strong></td>
<td><strong>Interactions</strong></td>
<td><strong>Notes:</strong> *$p &lt; 0.05$; **$p &lt; 0.01$; **<em>$p &lt; 0.001$</em></td>
</tr>
<tr>
<td>Gender</td>
<td>−0.072</td>
<td>−0.066</td>
<td>−0.063</td>
<td>−0.040</td>
</tr>
<tr>
<td>Age</td>
<td>−0.062</td>
<td>−0.035</td>
<td>−0.026</td>
<td>−0.037</td>
</tr>
<tr>
<td>Education</td>
<td>0.038</td>
<td>0.028</td>
<td>0.005</td>
<td>0.023</td>
</tr>
<tr>
<td>Work experience</td>
<td>0.086</td>
<td>0.071</td>
<td>0.036</td>
<td>0.068</td>
</tr>
<tr>
<td>Position</td>
<td>0.138</td>
<td>0.127</td>
<td>0.088</td>
<td>0.100</td>
</tr>
<tr>
<td><strong>IO (H1)</strong></td>
<td>−0.047</td>
<td><strong>−0.207</strong>*</td>
<td>−0.147*</td>
<td><strong>Table VI.</strong> Hierarchical regression analysis results</td>
</tr>
<tr>
<td><strong>ITO (H2)</strong></td>
<td>0.101</td>
<td>0.069</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td><strong>IPT (H3)</strong></td>
<td>0.221**</td>
<td>0.217***</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>JCAS (H4)</strong></td>
<td>0.494***</td>
<td>0.515***</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IPT × IO (H5a)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IPT × ITO (H5b)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>JCAS × IO (H6a)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>JCAS × ITO (H6b)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.030</td>
<td>0.038</td>
<td>0.372</td>
<td>0.470</td>
</tr>
<tr>
<td>Adj.$R^2$</td>
<td>0.002</td>
<td>−0.001</td>
<td>0.339</td>
<td>0.428</td>
</tr>
<tr>
<td>$ΔR^2$</td>
<td>0.030</td>
<td>0.008</td>
<td>0.354***</td>
<td>0.098***</td>
</tr>
</tbody>
</table>
6. Discussion, implications and future research

In this section, we summarize the key findings and potential contributions and implications of this study. The future research directions are also discussed.

6.1 Key findings

First, our results indicate that employees experience significant information overload that directly reduces their perceived job satisfaction. This violates managers’ initial intention of promoting mobile work in order to facilitate effective communication rather than become a potential barrier to knowledge sharing in the organization (Yuan et al., 2010). Furthermore, unlike traditional organizational IT, MICTs are embedded into individuals’ work and life. Thus, MICT users are more thoroughly addicted to the multiple sources of information than ever before. This result is consistent with previous literature that highlights the negative effects of information overload on individuals. However, it also is harder for both individuals and organizations to alleviate the negative effects of information overload in the context of mobile technology use.

Second, contrary to the negative characteristic of interruption overload in previous literature, interruption overload does not have a significant effect on individuals’ job satisfaction. In a traditional environment, the organization-related IT and IT used in personal life can be clearly separated. Therefore, the work tasks can be distinguished from individuals’ personal life. However, in the context of using MICTs, the boundaries between work and personal life are blurred and ambiguous. Thus, while individuals feel a certain level of interruption, the extent of the interruption might not exert a significant impact on individuals’ behavioral or attitudinal change to their work. This finding is also consistent with Ou and Davison (2011) who proposed that work interruptions caused by the use of instant messengers do not significantly affect group outcomes.

Furthermore, according to Addas and Pinsonneault (2015), IT interruptions may have both positive and negative effects. Furthermore, an individual’s work task boundaries also have a direct impact on how one perceives the nature and consequences of one’s interruptions. In the mobile technology use and multitasking context, employees’ task boundaries are blurred, with the result that they are unaware of interruption effects.

The above findings are interesting. Although individuals perceived increasing levels of cognitive load when using MICTs, these cognitive load effects do not significantly influence employees’ job satisfaction. This indicates that individual employees would prefer to enjoy the relatively ad hoc working style that is shaped by MICTs, and have more tolerance for the corresponding increasing levels of cognitive load and interruption caused by the use of MICTs. Furthermore, some prior studies have also shown that MICTs play a crucial role in individuals’ lives: these individuals form an intimate relationship with their mobile devices and may have already become accustomed to the continuous high level of cognitive load (Yoo, 2010; Bødker et al., 2014). In this way, the perceived cognitive load may not significantly negatively influence individuals’ perceived level of job satisfaction. This result is also consistent with some studies in the sense that the technostress caused by mobile technologies does not exert significant impacts on Chinese workers’ job satisfaction (Tu et al., 2005).

Third, the two proposed problem-focused coping factors (i.e. information processing timeliness and job control assistant support) significantly facilitate individuals’ work by supporting their efficiency and effectiveness. The designers of MICTs have paid increasing attention to user experience (Albert and Tullis, 2013; Olsson et al., 2013). The enhanced level of convenience, ease of use and usefulness of MICTs enable employees to handle complex work situations more flexibly, which subsequently increases their job satisfaction. Furthermore, the moderation effects of these two coping factors are also noteworthy. These two factors moderate the technology overload effects in different directions, but according
to our results, $H6a$ (moderating effects of job control assistant support on the information overload–job satisfaction relationship) is not significant. This suggests that the job control assistant support functions of MICTs are insufficient to help employees effectively reduce the perceived negative effects of information overload on job satisfaction. As illustrated in Section 2.1, information overload refers to both volume and time issues. This also indicates that although MICTs are used to facilitate efficient information processing, the large amount of information is still the main source of stress in the workplace. Thus, practicing managers should provide more effective business process control policies and related tools to decrease individual perceived overload levels.

Fourth, as shown in Table VI, the two key technology overload factors together explained only 3.8 percent of the variance of job satisfaction in Model 2. However, when we put the two problem-focused coping strategies (i.e. information processing timeliness and job control assistant support) into the model, we can see that these factors explained 37.2 percent of the variance of job satisfaction. The change in $R^2$ is 33.4 percent and is significant at the level of $p < 0.001$. This indicates the important role of MICTs with respect to reducing individual perceived negative impacts and increasing employees’ job satisfaction. This also helps us understand why users complain about the disturbance of MICTs but still continuously use them: MICTs give employees the sense that they are more in control of their daily work tasks. In organizations, the introduction of information technologies did increase employees’ work burden and make tasks more complex. However, MICTs provide a more flexible solution for employees to handle increasingly complex tasks and enable multitasking. This finding is supported by the principle of behavioral consistency theory (Wernimont and Campbell, 1968). That is, individuals who were previously successful at handling a task should enhance their self-efficacy and technical skills in that realm, increasing the probability of repeating that behavior.

6.2 Contributions and implications
With the widespread use of IT in organizations, the phenomenon of technology overload has aroused increasing attention from both practitioners and academics. In this study, we contribute to the literature by first distinguishing two concepts, information overload and interruption overload, and then integrating them into a cohesive framework of technology overload. The corresponding instruments have also been created and validated to facilitate researchers to further investigate the phenomenon of technology overload. As claimed by Karr-Wisniewski and Lu (2010), the concept of technology overload gives researchers a new problem to address. This study is in line with this stream of research and investigates the phenomenon of technology overload from a new perspective, in order to inspire more research in this field.

Furthermore, in the technostress literature, scholars have proposed the concept of “techno-overload” as an important component of technostress creators (Ayyagari et al., 2011; Ragu-Nathan et al., 2008). Rather than treating techno-overload as one dimension of technostress, we operated technology overload as an independent construct including two sub-components: information overload and interruption overload. Based on the CMUA, we investigated the user appraisal process while facing technology overload in the context of MICT usage at work. In addition, we identify information processing timeliness and job control assistant support as two moderators in the relationship between technology overload and job satisfaction. So the direct impacts of technology overload on employees’ job satisfaction have also been examined, according to the CMUA and cognitive load theory. Therefore, we provided a comprehensive conceptual model to understand the overload effects experienced by knowledge workers in modern workplace.

Our findings also provide useful insights for practitioners. The two identified technology overload factors can be treated as the indexes to examine the degree of employees’
perceived overload. This is useful for managers who wish to have a better understanding of the work situation of individuals, which may usefully contribute to a re-organization of their tasks. Our study has empirically demonstrated that the two identified coping strategies can significantly increase individuals' job satisfaction. In essence, these two MICT-supported functions actually improve individuals' work efficiency and effectiveness. Therefore, managers may invest more resources to provide available mobile functions of job control assistant support for employees, reducing employees' perceived overload. Finally, managers should pay attention to the moderating effects of the two coping factors. That is, the practitioners could achieve a better control of the technology overload effects by utilizing these MICT functions.

6.3 Limitations and future research
In this study, the development of information processing timeliness, as an important construct, should be further enhanced. Although the predictive validity and reliability are considered reasonable in this study, the limitations of a two-item scale for this construct means that there is room for improvement. Many studies have explored the phenomenon of information overload. However, the interruption overload has aroused little attention. It is worthwhile to investigate the interactions among interruption events, primary tasks and employee reactions in future work, so that more fluent workflow design implications can be drawn. We note that our survey participants work in the internet industry, where MICTs have been widely adopted. Therefore, although individuals have experienced high levels of cognitive load stress, the impact of these negative effects on individuals' job satisfaction may be not significant. Future studies should involve employees from different industries, so that comparisons can be made, and more specific suggestions can also be provided for practitioners. It would be interesting and valuable to include more individual characteristics or psychological factors in future studies. Furthermore, we considered a comprehensive range of mobile technologies in this study. However, in different contexts, employees may use different kinds of technologies to help them complete tasks. Different MICTs or social media as well as diverse system features may have diverse impacts on individuals. Future work can explore the context-dependent negative effects of MICTs, so that more specific implications and suggestions can be drawn to promote the further improvement of MICTs development and application in the workplace.

7. Conclusion
The dark side of IT has aroused increasing attention in recent years (Tarafdar et al., 2015). Some research has also been conducted to examine the increasing significant negative effects of social media (Mäntymäki and Islam, 2016; Fox and Moreland, 2015). However, the negative effects resulting from the use of MICTs (including various mobile social media applications and devices) have been paid little attention by academics. Specifically, with the prevalence of massive amounts of information, humans are experiencing increasing levels of overload. With the increasingly intimate relationship between humans and MICTs, the technology overload effects may be reflected in new forms, such as techno-dependency. However, few studies have been conducted to examine this phenomenon. The context-dependent technology overload effects have also not been identified or illustrated. Responding to the call of this special issue, this study constitutes the starting point for investigations into the phenomenon of technology overload. Future research needs to consider the different dimensions of technology overload and their corresponding influences on both individual performance and organizational behavior in different industry backgrounds.
References


Brumberg, R. (2018), “Study: workers waste 32 days per year juggling tech, app overload”, available at: www.prdaily.com/Marketing/Articles/1ad307c6-51c7-4494-ad0a-361019074a05.aspx (accessed April 8, 2018).


## Appendix 1

<table>
<thead>
<tr>
<th>Code</th>
<th>Constructs and measurements</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>IO</td>
<td>Information overload, scale: strongly disagree (1) to strongly agree (7)</td>
<td>Adapted from Eppler and Mengis (2004)</td>
</tr>
<tr>
<td>IO_1</td>
<td>When I use MICTs, the amount of information I have to deal with exceeds my capability to process it</td>
<td></td>
</tr>
<tr>
<td>IO_2</td>
<td>When I use MICTs, the increased amount of information makes me feel overloaded</td>
<td></td>
</tr>
<tr>
<td>IO_3</td>
<td>When I use MICTs, the speed of information processing is beyond my ability</td>
<td></td>
</tr>
<tr>
<td>ITO</td>
<td>Interruption overload, scale: strongly disagree (1) to strongly agree (7)</td>
<td>Adapted from Ou and Davison (2011)</td>
</tr>
<tr>
<td>ITO_1</td>
<td>I feel overload because my current work is always interrupted by MICTs</td>
<td></td>
</tr>
<tr>
<td>ITO_2</td>
<td>I feel overload because using MICTs always inhibits my concentration on work</td>
<td></td>
</tr>
<tr>
<td>ITO_3</td>
<td>I feel that MICTs are quite disturbing which makes me feel overloaded</td>
<td></td>
</tr>
<tr>
<td>JOBSA</td>
<td>Job satisfaction, scale: strongly disagree (1) to strongly agree (7)</td>
<td>Adapted from Locke (1976)</td>
</tr>
<tr>
<td>JOBSA_1</td>
<td>My job is enjoyable because I use MICTs</td>
<td></td>
</tr>
<tr>
<td>JOBSA_2</td>
<td>I enjoy cooperating with my colleagues by using MICTs</td>
<td></td>
</tr>
<tr>
<td>JOBSA_3</td>
<td>Generally speaking, by using MICTs I feel satisfied with my job</td>
<td></td>
</tr>
<tr>
<td>IPT</td>
<td>Information processing timeliness, scale: strongly disagree (1) to strongly agree (7)</td>
<td>Developed in this study based on Liang et al. (2007)</td>
</tr>
<tr>
<td>IPT_1</td>
<td>When I use MICTs, I need to deal with information more quickly</td>
<td></td>
</tr>
<tr>
<td>IPT_2</td>
<td>When I use MICTs, the way I handle information is constrained by tight time schedules</td>
<td></td>
</tr>
<tr>
<td>JCAS</td>
<td>Job control assistant support, scale: strongly disagree (1) to strongly agree (7)</td>
<td>Developed in this study based on Hung et al. (2011)</td>
</tr>
<tr>
<td>JCAS_1</td>
<td>Use of MICTs enables me to set (or get) advanced notices (or reminders) to control my work</td>
<td></td>
</tr>
<tr>
<td>JCAS_2</td>
<td>Use of MICTs enables me to easily manage work tasks</td>
<td></td>
</tr>
<tr>
<td>JCAS_3</td>
<td>Use of MICTs enables me to take full advantages of fragmented time, thus I feel workload is reduced</td>
<td></td>
</tr>
<tr>
<td>JCAS_4</td>
<td>Generally speaking, use of MICTs is helpful to me to improve my job performance</td>
<td></td>
</tr>
</tbody>
</table>

Table AI. Construct measures
## Appendix 2

### Table AII.

Demographics of Employees in the internet industry in China in 2016

<table>
<thead>
<tr>
<th>Categories</th>
<th>Items</th>
<th>%</th>
<th>Categories</th>
<th>Items</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16–20</td>
<td>2.19</td>
<td></td>
<td>Male</td>
<td>65.87</td>
<td></td>
</tr>
<tr>
<td>21–25</td>
<td>32.90</td>
<td></td>
<td>Female</td>
<td>34.13</td>
<td></td>
</tr>
<tr>
<td>26–30</td>
<td>29.72</td>
<td></td>
<td>Working experience (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31–35</td>
<td>21.14</td>
<td></td>
<td>&lt; 1</td>
<td>9.02</td>
<td></td>
</tr>
<tr>
<td>36–40</td>
<td>11.01</td>
<td></td>
<td>1–3</td>
<td>28.44</td>
<td></td>
</tr>
<tr>
<td>41–45</td>
<td>3.03</td>
<td></td>
<td>3–5</td>
<td>23.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5–7</td>
<td>19.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>7–10</td>
<td>13.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&gt; 10</td>
<td>6.08</td>
<td></td>
</tr>
</tbody>
</table>

**Source:** The White Paper of China Internet Workplace Eco (2016)

---

**Corresponding author**

Pengzhen Yin can be contacted at: yinpengzhen888@126.com

---

For instructions on how to order reprints of this article, please visit our website: [www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: permissions@emeraldinsight.com
Is my fear of missing out (FOMO) causing fatigue? Advertising, social media fatigue, and the implications for consumers and brands

Laura Frances Bright
Department of Strategic Communication, Texas Christian University, Fort Worth, Texas, USA, and

Kelty Logan
Department of Advertising, Public Relations and Media Design, University of Colorado Boulder, Boulder, Colorado, USA

Abstract

Purpose – Social media usage has become ubiquitous in our society – consumers are spending upwards of 20 percent of their media time on social sites interacting with friends, family and brands (Adler, 2016) and all of this usage is driving fatigue. The purpose of this paper is to examine how advertising factors such as attitude and intrusiveness impact social media fatigue as well as two consumer behavior factors, fear of missing out (FOMO) and privacy.

Design/methodology/approach – A 190-item questionnaire was developed and administered to an opt-in subject pool recruited for web-based research (i.e. online panel). A representative sample of 750 US social media users was recruited for the survey of which 518 respondents were valid and usable.

Findings – Results indicate that FOMO has the greatest impact on social media fatigue, not advertising factors as predicted. In addition, privacy concerns continue to plague consumers and should be monitored by advertisers.

Research limitations/implications – With regard to limitation, the survey contained a variety of self-reported measures that can tend to be under-reported, especially when it comes to social media usage as evidenced in a recent study (Adler, 2016).

Originality/value – This research undertook an investigation of consumer perceptions of social media advertising and how those relate to social media fatigue and psychological factors such as privacy and FOMO. In looking at these constructs, a clearer picture of how consumer perceptions of advertising impact levels of social media fatigue has emerged. In addition, the results provide a better understanding of FOMO, a psychological factor that significantly contributes to social media fatigue.

Keywords: Privacy, Advertising, Social networks, Attitudes, Information processing, Media fatigue

Paper type: Research paper

Introduction

Casting a cursory glance at today’s society, even the untrained eye can see media consumption in abundance. Consumers are glued to their smartphones in a pattern of habitual usage that provides them immediate access to news, social networks, pop culture, location-based information and much more. Driven by advances in technology and smart phone access, social media have infused themselves into our lives in an unprecedented manner. In fact, social networking leads all categories of digital media in engagement – including search, multimedia, radio and gaming – accounting for one out of five minutes spent online (comScore, 2015, p. 29). In all, 69 percent of US adults now use social media sites, a substantial increase from a decade ago when only 21 percent of US adults used these services (Pew Research Center, 2018). As a consequence, social media is becoming an increasingly important element of the advertising and promotions mix used by brands to communicate with their consumers. Over the course of 2018, US advertisers will spend nearly $25bn on social media advertising, and spending is
projected to reach $34bn by 2021 (eMarketer, 2018). With these substantial investments in mind, it is vital to know more about how consumers are interacting with brands in the social media environment and whether the abundance of brand information is contributing to social media fatigue.

There are growing indications that Millennials, the key 18-to-34-year-old consumer segment, are showing signs of “social media fatigue,” the tendency to back away from consumption when too much content is in their social media feeds. Indeed, in a recent article, Andrew Sullivan referred to this tendency as a “distraction sickness.” “My brain had never been so occupied so insistently by so many different subjects and in so public a way for so long,” he said of his own experience with distraction sickness (Sullivan, 2016). This feeling could also be linked to a concept known as “fear of missing out” (FOMO) which leads consumers to habitually consume social media content because they are afraid of missing out on important information or happenings (Przybylski et al., 2013).

Research has explored the appeal and marketing potential of social media from many perspectives. Shao (2009) provided a uses and gratifications analysis to determine the factors that increase the appeal of social media usage, in general. Other studies have investigated the use of social media data to predict consumer behavior (Kalampokis et al., 2013; Schoen et al., 2013). Research has also examined specific marketing opportunities, such as the effect of online brand communities on brand satisfaction (Zhang and Luo, 2016), the factors that influence the effectiveness of product placement in social media (Liu et al., 2015), and the influence of Facebook advertising with regard to purchase intent (Duffett, 2015).

In terms of social media fatigue, however, the research is limited. One study has ascertained several of the psychological factors that can contribute to social media fatigue, including a consumer’s level of confidence, perceived self-efficacy regarding social media use, perceived helpfulness of social media and privacy concerns with sharing data on social media sites (Bright et al., 2015). The authors suggest that consumers’ social media fatigue is largely attributable to internet-based privacy concerns and over-reliance upon social media due to its perceived helpfulness. Articles in the trade press, on the other hand, suggest that advertising intrusiveness may contribute to social media fatigue. A recent Harris Poll (2016) stated that 74 percent of young people, ages 16–36, object to being targeted by marketers in their social media feeds and 56 percent of young people stated that they reduced or terminated their use of specific social media sites due to advertising in their feeds. To date, there is no academic research to explain the relative effects of consumers’ perceptions of social media advertising compared to the psychological factors that ultimately cause social media fatigue. To further this understanding, this study examines consumers’ perceptions of social media advertising, whether those perceptions contribute to social media fatigue, and how they relate to the psychological factors that affect social media fatigue, such as privacy concerns and FOMO.

**Literature review**

**Current state of social media**

In all, 69 percent of all adults and 88 percent of all young adults aged 18–29 are on social media (Pew Research Center, 2018). Social networking is the fourth most common internet activity (Kittenger et al., 2012). The majority of social media users utilize these sites primarily for social purposes, with two-thirds of users recognizing that “staying in touch with current friends and family members is a major reason they use these sites” and half saying the same about reconnecting with old friends (Smith, 2011).

A study attempting to determine whether or not Millennials are addicted to social media revealed that subjects reported experiencing salience, tolerance, intrapsychic conflict and relapse associated with their social media use (Cabral, 2011). According to Cabral (2011), the association of these aspects of addiction with social media use shows that Millennials
prioritize social media above other activities and need to use social media more often to achieve satisfaction. The subjects of this study acknowledged that they needed to cut back on their social media usage but admitted how challenging it would be. These findings show how such habitual use could lead to feelings of fatigue with social media content, platforms and networks. They also show how FOMO contributes to the difficulty of moderating usage as consumers increasingly rely upon social media for the most up-to-date content from friends and family and special offers from their favorite brands.

With the permeation of technology into every aspect of Millennials’ lives today, social media has truly become essential to Generation Y. In his book *The Shallows*, Nicholas Carr (2010) argued that internet technologies now “guide people’s behavior and shape their perceptions” and that because of this, today’s society exists in a world of interruptions. These interruptions are no accident, however. They exist because of people’s compulsion to be constantly connected. Carr proposes that these disruptions are changing the way people think and process information and that the internet is, in part, responsible for an extensive loss of ability to focus. Given the widespread and constant reliance on internet technology, people are no longer able to avoid distractions. Furthermore, it appears they are actually inviting those distractions. Interruptions have become the norm and the ability to focus has become more challenging (Carr, 2010; Mott, 2013). Due to the large investments brands are making in social media, it is vital to tease apart how this environment of distraction and fatigue is impacted by the presence of brand messages.

**Social media fatigue**

Social media fatigue is defined as “social media users’ tendency to back away from social media usage when they become overwhelmed with too many sites, too many pieces of content, too many friends and contacts, and too much time spent keeping up with these connections” (Technopedia, 2011). Social media fatigue can also be linked to concerns about privacy and boredom among social media users.

According to a report issued by Goasduff and Pettey (2011), the social media market is beginning to show signs of maturity as some users in certain segments show signs of social media fatigue. The company surveyed 6,285 users between the ages 13 and 74 in 11 developed and developing markets from December 2010 and January 2011. Respondents were asked about their use and opinions regarding social media. Of those surveyed, 24 percent said they were using their favorite social media site less than when they first signed up. Those who admitted to less social media use were represented by consumer segments that exhibit a practical view of technology (Goasduff and Pettey, 2011). Brian Blau, researcher at Gartner Inc., reported “The trend shows some social media fatigue among Early Adopters, and the fact that 31% of Aspirers [younger, more mobile, brand-conscious consumers] indicated that they were getting bored with their social network is a situation that social media providers should monitor, as they will need to innovate and diversify to keep consumer attention” (Goasduff and Pettey, 2011, p. 1). When asked why they were tiring of social media, the respondents indicated that online privacy was their greatest concern. However, teenagers and those in their 20s were significantly more likely to report increased usage of social media. In fact, 37 percent of those in younger age groups and more tech-savvy segments said they used their favorite social media site more (Goasduff and Pettey, 2011). This paradox faced by Millennials (feeling fatigued yet increasing usage) could potentially be driven by FOMO on current and breaking content from friends, family, brands and news outlets.

In some recent studies, researchers have analyzed how this new media environment may be compromising learning and attention and, ultimately, impacting retention of advertising-related information as well. One such study concluded that individuals pay less attention to marketing messages when faced with distractions (Lang, 2000) and retain less information overall. Other studies have shown that recall can be inhibited (Nasco and
Biocca et al. (2007) found that overwhelming amounts of information make it harder to comprehend complex messages, and Palfrey and Gasser (2008) found that multitasking and information overload can cause people to use sensory filters to cope with the sheer volume of information, making it impossible to pay attention to most messages (Hill and Moran, 2011). Furthermore, psychologists have linked multitasking (including text messaging and instant messaging) with self-reported symptoms of depression and social anxiety among college undergraduates (Becker et al., 2013). Yet, little of this research has dealt specifically with social media as a context nor looked at how advertising messages from brands might be contributing to the feelings reported above.

Social media fatigue can come from interaction with advertising and brands as well. Studies show that consumers seek discounts and consumer reviews (eMarketer, 2015) while brands want to use social media to provide information about new products and get opinions about their performance (eMarketer, 2014). The disparity between what brands are delivering to consumers and what consumers want from brands adds yet another layer of complexity to brand communication in the social media environment. As such, brands must figure out how to balance their communication efforts with consumers so that they are not contributing to social media fatigue in this oversaturated marketplace.

Attitude toward social media advertising
As a construct, attitude toward advertising involves both general and personal assessment. The area of general assessment tends to regard advertising as an institution while the personal assessment reflects experience with actual ads. MacKenzie and Lutz (1985) suggested that a consumer’s response to an ad is mediated by the consumer’s prior knowledge and experience regarding advertising in general. Attitude toward advertising in general is also related to a consumer’s predisposition to attend or avoid a specific ad. Greyser (1973) stated that public dislike of advertising leads to inattention. Mittal (1994) indicated that a significant relationship exists between negative attitudes toward advertising and advertising avoidance. Moreover, because advertising avoidance opposes attendance to the advertising, the behavior is related to reduced advertising effectiveness (Shavitt et al., 2004). In the case of social media, negative attitudes toward social media advertising that lead to avoidance of advertising in the medium should correspond to avoidance of social media use. It is reasonable to assume, therefore, that the more negative consumers’ attitudes toward social media advertising, the more likely they would experience social media fatigue:

\[ H1. \text{ Attitude toward social media advertising relates negatively to social media fatigue in consumers.} \]

Attitude toward brands
Research has indicated that brand attitudes reflect perceptions regarding brand attributes and brand advertising (Mitchell and Olson, 1981). Consumers’ general attitudes toward brands, therefore, are assumed to incorporate an evaluative element that indicates a positive or negative assessment (Eagly and Chaiken, 1993) that is relatively stable and enduring over time.

Furthermore, research also indicates that a consumer’s attitude toward a brand, “energizes and directs behavior” in regard to the brand (Eagly and Chaiken, 1993, p. 7), such as purchase intent. More recently, it has been suggested that while attitude strength predicts purchase behavior, brand attachment – which is defined as the strength of the bond connecting the consumer with a brand – provides a more precise measure of the attitude valence and, therefore provides a more accurate measure (Park et al., 2010). Chu et al. (2015) found that brand attachment drove consumer intention to follow brands on social media, specifically Twitter. Because industry research indicates that those who follow brands on social media are most interested in information regarding sales, discounts and coupons, it is
likely that those who follow brands will have positive attitudes toward social media advertising because they receive a benefit from following their favorite brands online (i.e. saving money with coupons or being the first to be offered new products):

**H2. Consumers’ attitude toward brands relates positively to their attitude toward social media advertising.**

*Advertising intrusiveness in social media*

Li et al. (2002) approached the need for a media-agnostic measure of communication problems by developing a scale to measure advertising intrusiveness. Advertising intrusiveness is defined as "a psychological reaction to ads that interfere with a consumer’s ongoing cognitive processes" (p. 39). The purpose of the advertising intrusiveness scale is to measure the source of the irritation rather than the negative emotions evoked by advertising or the various means of ad avoidance. Their research confirmed, however, that the scale correlates with measures of irritation and behavioral advertising avoidance.

Advertising avoidance measures were limited to observable behavior and, therefore, did not include measures of cognitive or mechanical avoidance. Rather, the advertising intrusiveness measure focuses on the negative behavioral consequences of negative advertising perceptions. In other words, the greater consumers’ perceptions of social media advertising intrusiveness, the more likely they would choose to avoid the advertising and, hence, experience social media fatigue:

**H3. Consumers’ perceptions of social media advertising intrusiveness relate positively to the degree of social media fatigue they experience.**

*Concerns with privacy*

As social media use grows and the options expand, the issue of privacy becomes ever more important. Facebook continues to have a liberal viewpoint on privacy and Google recently announced major changes to privacy among its properties. In general, the transparent interaction between user and site raises concerns about privacy online (Karahasanovic et al., 2009). Indeed, research from Goasduff and Pettey (2011) shows that privacy is a major concern with users. Thus, it is expected that those with higher privacy concerns will also have social media fatigue due to the “threshold beyond which social contact becomes irritating for all parties” (Schwarz, 1968, p. 741). Additionally, there are continued concerns about what social media sites do with the information collected. If, as the Harris Poll (2016) suggested, consumers view social media advertising as intrusive, it is likely that negative attitudes toward the advertising contributes to their privacy concerns:

**H4. Consumers’ attitude toward social media advertising relates negatively to the degree of privacy concerns they experience.**

*Fear of missing out*

Consumers have a seemingly infinite amount of information available in their feeds yet a finite amount of time to process the information. Given the paradox of social media usage, some can experience a concept known as FOMO if they are not constantly up to date on new information from their network. FOMO is defined as “a pervasive apprehension that others might be having rewarding experiences from which one is absent” (Przybylski et al., 2013, p. 1841). Indeed, a recent study conducted among adolescents found that FOMO mediates the relationship between psychopathology and the negative consequences related to mobile phone usage. Specifically, it was found that depression triggered overuse of social media among female adolescents, while anxiety served as a trigger for male adolescents
(Oberst et al., 2017). Building on this literature, FOMO could be related to both experiences with friends and family as well as marketplace experiences such as special offers on products or services. Consumers will often opt in to receiving branded content on social media by liking a brand page. This allows them to receive special offers and discounts so that they do not miss out on any deals. While this satisfies the consumers’ needs for information and reduces their FOMO, it could also lead to social media fatigue if the brand is distributing large amounts of content via social media:

- H5a. FOMO relates positively to attitude toward brands.
- H5b. FOMO relates positively to attitude toward social media advertising.
- H5c. FOMO relates positively to attitude toward social media.
- H6. FOMO relates positively to social media fatigue.

Methodology

Design

In all, 46 items were pre-tested on a small group of academic professionals to ensure clarity of the constructs under investigation. Once tested, the items were included in a 190-item online survey that incorporated additional items to pre-screen respondents, obtain demographic information and create composite variables for two additional research projects. The survey was administered using Qualtrics, a web-based survey management system, and administered to an opt-in subject pool recruited for web-based research (i.e. online panel). A representative sample of US social media users was recruited for the survey. Data were gathered from a total of 750 participants over a seven-day period in August 2016 to ensure an even distribution of respondents on each day of the week (i.e. weekdays and weekend days).

Sample

The final sample consisted of 518 current US social media users, 46.9 percent (n = 243) male and 53.1 percent (n = 275) female, who indicated that they follow brands on their social media accounts. Among this sample, respondents ranged in age from 18 to 65 with the majority of respondents between the ages of 18 and 40 (55.2 percent, n = 286). With regard to race or ethnicity, 68.0 percent (n = 352) were Caucasian, 12.4 percent (N = 64) were African–American, 5.2 percent (n = 27) were Asian–American, 9.5 percent (N = 49) were Hispanic American, 1.2 percent (n = 6) were American–Indian, 2.1 percent (n = 11) were multiracial, 0.4 percent (n = 2) were international and 1.4 percent (n = 7) preferred not to indicate their race or ethnicity. The majority of the sample had an annual household income level of $50,000 or above (60 percent, n = 311) with the remaining sample being evenly distributed among an annual household income of $40,000–$49,999 (8.5 percent, n = 44), $20,000–$39,999 (19.9 percent, n = 103), and below $20,000 (11.6 percent, n = 60). In terms of education level, 22.6 percent (n = 117) of the sample had undertaken some college, 17 percent (n = 88) were high school graduates, 22.8 percent (n = 118) received a bachelor’s degree, 17.8 percent (n = 92) had a graduate degree, 12.2 percent (n = 63) completed an associate’s degree, 2.9 percent (N = 15) had completed some high school and 4.8 percent (n = 25) had completed some graduate school.

Design

Measures

Established scales were used to measure the seven variables. Cronbach’s $\alpha$ reliabilities for the scores obtained from each of the scales were interpreted in accordance with George and Mallery (2003). Specifically they suggest that reliabilities greater than 0.9 are “excellent,”
greater than 0.8 are “good,” greater than 0.7 are “acceptable,” greater than 0.6 are “questionable,” greater than 0.5 are “poor” and less than 0.5 are “unacceptable.” The individual items for each scale are provided in Table AI.

Independent variables
To measure privacy concerns regarding social media, respondents were asked to indicate their agreement with a series of statements on a four-item, seven-point scale (1 = “Strongly Agree” and 7 = “Strongly Disagree”) adapted from a Gartner study (Goasduff and Pettey, 2011). The responses were averaged to form a scale (α = 0.65, M = 5.12, SD = 1.27).

FOMO was measured using a ten-item, five-point scale (1 = “Not at all true of me” and 5 = “Extremely true of me”) that was adapted from Przybylski et al. (2013). The responses demonstrated excellent reliability and were averaged to form a scale (α = 0.94, M = 2.31, SD = 1.02).

Attitude toward social media was measured using a six-item, five-point scale (1 = “Strongly Agree” and 5 = “Strongly Disagree”) adapted from Bearden et al. (2011). The responses demonstrated excellent reliability and were averaged to form a scale (α = 0.92, M = 3.57, SD = 0.95).

Attitude toward brands was measured using an eight-item, seven-point scale (1 = “Strongly Agree” and 7 = “Strongly Disagree”) adapted from Bearden et al. (2011). The responses demonstrated excellent reliability and were averaged to form a scale (α = 0.97, M = 4.17, SD = 1.55).

Attitude toward social media advertising was measured using an established, four-item, three-point semantic differential scale (Elliott and Speck, 1998) that was adapted to social media advertising. Respondents were asked to indicate, for example, how interesting they found social media advertising (1 = Interesting; 3 = Not Interesting). The responses were reverse coded to match the directionality of the other survey responses (negative to positive) to simplify interpretation of the analytical results. The responses demonstrated good reliability and were averaged to form a scale (α = 0.89, M = 1.86, SD = 0.61).

Social media advertising intrusiveness was measured using an established seven-item, seven-point Likert-type (1 = “Strongly Agree” and 7 = “Strongly Disagree”) scale by Li et al. (2002). Items were adapted to discuss social media advertising specifically. The responses demonstrated excellent reliability and were averaged to form a scale (α = 0.96, M = 5.56, SD = 1.70).

Dependent variable
Respondents’ level of social media fatigue was measured using a seven-item, seven-point scale (1 = “Strongly Agree” and 7 = “Strongly Disagree”) scale adapted from a study conducted by Goasduff and Pettey (2011). The responses demonstrated excellent reliability and were averaged to form a scale (α = 0.87, M = 4.09, SD = 1.16).

Results
SPSS statistical software was used for all statistical analyses. Multiple regression analysis was used to assess the net effect of each variable upon social media fatigue. Pearson correlations were used to assess the strength of the relationships among the key variables.

Regression analysis
Multiple regressions were utilized to analyze the net effects of each of the six independent variables (privacy concerns, FOMO, attitude toward social media, attitude toward brands, attitude toward social media advertising and social media advertising intrusiveness) upon social media fatigue.
Regarding social media fatigue, the six predictors explained 40 percent of the variance, 
\( R^2 = 0.40 \), adj. \( R^2 = 0.39 \), F(6, 358) = 40.11, \( p < 0.001 \). Privacy concerns (\( \beta = 0.35, p < 0.001 \)), 
FOMO (\( \beta = 0.44, p < 0.001 \)), attitude toward social media (\( \beta = -0.17, p < 0.001 \)) and social 
media advertising intrusiveness (\( \beta = 0.10, p < 0.05 \)) each significantly predicted social 
media fatigue.

Table I displays the unstandardized regression coefficients (\( B \)), standard error, standardized 
regression coefficients (\( \beta \)), t-test statistic and significance level (\( p \)) for each variable.

### Correlation Analysis

Table II shows the relationship among the key independent variables and social media 
fatigue. While all of the relationships are significant, variance inflation factor (VIF) tests 
revealed that the multicollinearity issues were not a problem. Specifically, the VIFs were less 
than “2” in the worst cases (Hair et al., 1995). It should also be noted that, compared to the 
other variables, the correlation coefficients are much smaller when considering the 
relationships among social media fatigue, attitude toward social media, attitude toward 
social media advertising and social media advertising intrusiveness.

Attitude toward social media advertising is positively, significantly related to social 
media fatigue, \( r = 0.16 (p < 0.01) \). This indicates that the more positive consumers’ attitude 
toward social media advertising, the more social media fatigue they will experience. This 
finding does not support \( H1 \).

Attitude toward brands is positively, significantly related to attitude toward social media 
advertising, \( r = 0.48 (p < 0.01) \). This finding indicates that consumers with a positive 
attitude toward brands will more likely have a positive attitude toward social media 
advertising. This finding supports \( H2 \).

Social media advertising intrusiveness is positively, significantly related to social media 
fatigue, \( r = 0.13 (p < 0.01) \). This result indicates that the more intrusive a consumer finds

<table>
<thead>
<tr>
<th>Predictor variable</th>
<th>B</th>
<th>SE B</th>
<th>( \beta )</th>
<th>t</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social media privacy</td>
<td>0.38</td>
<td>0.05</td>
<td>0.35</td>
<td>7.80</td>
<td>0.000</td>
</tr>
<tr>
<td>Fear of missing out (FOMO)</td>
<td>0.51</td>
<td>0.06</td>
<td>0.44</td>
<td>9.06</td>
<td>0.000</td>
</tr>
<tr>
<td>Attitude toward brands</td>
<td>0.07</td>
<td>0.04</td>
<td>0.09</td>
<td>1.68</td>
<td>0.095</td>
</tr>
<tr>
<td>Attitude toward social media</td>
<td>-0.22</td>
<td>0.07</td>
<td>-0.17</td>
<td>-3.41</td>
<td>0.001</td>
</tr>
<tr>
<td>Attitude toward social media advertising</td>
<td>0.09</td>
<td>0.10</td>
<td>0.05</td>
<td>0.90</td>
<td>0.367</td>
</tr>
<tr>
<td>Social media advertising intrusiveness</td>
<td>0.07</td>
<td>0.04</td>
<td>0.10</td>
<td>2.11</td>
<td>0.036</td>
</tr>
</tbody>
</table>

**Notes:** \( B \), unstandardized \( \beta \) coefficient; SE \( B \), standard error; \( \beta \), standardized \( \beta \) coefficient; \( t \), t-test statistic; \( p \), significance value

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Social media fatigue</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Social media privacy</td>
<td>0.43**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Fear of missing out (FOMO)</td>
<td>0.48**</td>
<td>0.18**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Attitude toward brands</td>
<td>0.35**</td>
<td>0.30**</td>
<td>0.50**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Attitude toward social media</td>
<td>0.15**</td>
<td>0.23**</td>
<td>0.43**</td>
<td>0.49**</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Attitude toward social media advertising</td>
<td>0.16**</td>
<td>0.10*</td>
<td>0.33**</td>
<td>0.48**</td>
<td>0.32**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>7. Social media advertising intrusiveness</td>
<td>0.13**</td>
<td>0.11*</td>
<td>-0.07</td>
<td>-0.22**</td>
<td>-0.17**</td>
<td>-0.50**</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**Notes:** \( n = 518 \). *\( p < 0.05 \), **\( p < 0.01 \)
social media advertising, the more likely the consumer will experience social media fatigue. This finding supports $H3$.

Attitude toward social media advertising is positively, significantly related to privacy concerns, $r = 0.10$ ($p < 0.05$). As previously mentioned, the correlation coefficient between the two variables is much smaller than the correlation coefficients among privacy concerns, attitude toward social media and attitude toward brands. The relationship does suggest, however, that consumers' perceptions of social media advertising relate to their privacy concerns regarding social media. This finding does not support $H4$.

FOMO is positively, significantly related to attitude toward brands, $r = 0.50$ ($p < 0.01$), attitude toward social media advertising, $r = 0.33$ ($p < 0.01$) and attitude toward social media, $r = 0.43$ ($p < 0.01$). These findings indicate that the greater a consumers' need for information, the more positive their attitude toward social media will be including attitude toward brands and social media advertising. These findings support $H5a–H5c$.

Social media fatigue is also positively, significantly related to FOMO, $r = 0.48$ ($p < 0.01$), indicating that the more an individual relies upon social media to stay informed, the more likely the individual will experience social media fatigue. This finding supports $H6$.

Conclusions

Discussion

This research undertook an investigation of consumer perceptions of social media advertising and how those relate to social media fatigue. Specifically, psychological factors such as privacy and FOMO were examined among consumers who follow brands on social media sites. In looking at these constructs, a clearer picture of how consumer perceptions of advertising impact levels of social media fatigue has emerged. In addition, the results provide a better understanding of FOMO, a psychological factor that significantly contributes to social media fatigue. The results show that while social media presents a “high efficiency, low friction path for those who are oriented toward a continual connection with what is going on,” it can also lead to feelings of fatigue (Przybylski et al., 2013, p. 1842). The FOMO paradox rings true with these respondents: consumers want more information and have a positive attitude toward following brands on social media sites but once they start receiving all the information, they become overwhelmed. In other words, consumers enjoy subscribing to information feeds until they have to wade through the content on a daily basis.

Previous research has established that privacy concerns are among the primary drivers of social media fatigue in US social media users (Bright et al., 2015). As expected, the study indicates that concerns about social media privacy are a significant predictor of social media fatigue. To see if other factors could further explain the impact of privacy on social media fatigue, this study looked at the relationship of privacy to various advertising-related constructs. Contrary to expectations, privacy concerns were not negatively related to any of the social media advertising variables (attitude toward advertising, advertising intrusiveness or attitude toward brands). The findings indicate that, while those who follow brands on social media report higher levels of social media fatigue, attitude toward social media advertising and attitude toward brands are not significant predictors of social media fatigue. Perhaps this is due to the opt-in nature of social media advertising. Consumers must “like” or “follow” a brand and actively give them permission to deliver marketing messages to their feed. In doing so, consumers are agreeing to share a certain level of information in exchange for exclusive content, thus overtly and voluntarily forfeiting their privacy concerns. Consumer perceptions of social media advertising intrusiveness, however, are negatively related to attitude toward brands, social media in general, and social media advertising. This suggests that some consumers opt out of social media advertising because it is annoying or intrusive rather than a privacy concern.
Implications to research

The most interesting finding in this study is that consumers experience conflicting wants regarding their need for information and their desire to avoid information overload. In fact, the research indicates that FOMO is directly related to social media fatigue. As consumers rely more and more on digital media to engage with brands, they can experience FOMO if they do not receive up-to-date content and offers from the brands they follow on social media. However, consistently having this information delivered via their social media feeds can also lead to fatigue, creating a paradox for advertisers. The degree that consumers use social media to remain informed about their friends and family is positively related to the likelihood of experiencing social media fatigue. FOMO is, in fact, the strongest predictor of social media fatigue. Unlike their attitudes toward social media advertising, consumers have not suspended their privacy concerns regarding the use of social media sites in general. Consumers appear to regard social media sites with suspicion and their privacy concerns increase in tandem with their need to consult social media sites for the latest information. The more they consult the sites, the more likely they are to experience social media fatigue. Given that consumers have indicated a willingness to discontinue use of social media sites when they become fatigued, the privacy concerns they are experiencing will eventually affect advertisers. Therefore, advertisers should become involved in the issue of social media users’ privacy protection.

However, consistent with prior research, privacy concerns are a significant predictor of social media fatigue, suggesting that participation in social media generates a general concern regarding consumer privacy. Given the frequent changes in policies and updates on various social media platforms, consumers may experience uncertainty regarding the amount of privacy protection provided by the sites and, ultimately, their worries will continue to contribute to social media fatigue. Advertisers and brands would be best served by being transparent about their data usage policies on social media sites.

Implications to practice

This study shows that advertisers face a unique challenge when developing strategies for social media. They must satiate a consumer’s FOMO by providing exclusive content and up-to-date brand information; however, they need to be at a cadence that will not induce social media fatigue or contribute to a change in consumers attitudes toward social media sites. If fatigue is induced, advertisers risk losing consumers (via unsubscribing) due to the intrusive content. Cundari (2015) pointed out the importance of planning out this cadence properly so that consumers will have the mental energy to interact with branded content on social media sites.

Limitations

As with any academic research endeavor, this study had a variety of limitations to report. First, the survey contained approximately 200 questions and could possibly have caused survey fatigue among the sample – this was evidenced by several data sets that were timed out and thus removed before analysis. The survey also contained a variety of self-reported measures including social media usage, general media usage, individual psychological differences and attitude toward brands and social media content. These measures can tend to be under-reported, especially when it comes to social media usage as evidenced in a recent study by Adler (2016). To address these limitations, future data collections will focus on newly developed scales to measure social media engagement and time spent via behavioral data collection.

Future research

Future research should address specific measures that would assuage consumers’ privacy concerns. If, in fact, advertisers have alleviated concerns by using “opt-in” measures,
perhaps the social media sites should provide similar options to ensure greater user satisfaction and, ultimately, loyalty. In addition, future research could investigate how consumers respond to different types of content and what drives fatigue the most. It is possible, for example, that non-personalized advertising and branded content does not alarm social media users but does raise concerns regarding how they have been targeted. In other words, social media users may be blaming the sites rather than the advertisers. Finally, future research could continue to examine how FOMO guides the consumer/brand relationship in the social media context.

The results of this study show that social media users exhibit a positive attitude toward brands and that social media channels provide a cost-effective and efficient vehicle for delivering that information. It is suggested that advertisers target their communication to specific consumer segments rather than a broad audience in an effort to reduce advertising over-exposure and subsequent social media fatigue among consumers.

References


Cundari, A. (2015), Customer-Centric Marketing: Build Relationships, Create Advocates, and Influence your Consumers, Amazon Digital Services, LLC, Seattle, WA.


**Further reading**


(The Appendix follows overleaf.)
Appendix

Social media fatigue
Smfatigue1 I am likely to receive too much information when I am searching
Smfatigue2 I am frequently overwhelmed by the amount of information available on social media
Smfatigue3 I find that social media sites do not have enough detail to quickly find the information that I need
Smfatigue4 The amount of information available on social media sites makes me feel tense
Smfatigue5 The amount of information available on social media sites makes me feel overwhelmed
Smfatigue6 When searching for information on social media sites, I frequently just give up because I get too overwhelmed
Smfatigue7 I am confident in my ability to deal with large amounts of information on social media sites

Social media privacy concerns
Smprivacy1 I am concerned about my privacy on social networks
Smprivacy2 I believe that my personal information can easily be used by marketers
Smprivacy3 I feel that I have to give too much information to social networks
Smprivacy4 I am confident that setting my privacy settings keeps my information on social networks private

Fear of missing out (FOMO)
Fomo1 I fear others have more rewarding experiences than me
Fomo2 I fear my friends have more rewarding experiences than me
Fomo3 I get worried when I find out my friends are having fun without me
Fomo4 I get anxious when I do not know what my friends are up to
Fomo5 Sometimes I wonder if I spend too much time keeping up with what is going on
Fomo6 It bothers me when I miss an opportunity to meet up with friends
Fomo7 When I have a good time, it is important for me to share details online
Fomo8 When I miss out on a planned get-together it bothers me
Fomo9 When I go on vacation, I continue to keep tabs on what my friends are doing

Attitude toward social media
Smattitude1 Social media is part of my everyday activity
Smattitude2 I am proud to tell people I am on social media
Smattitude3 Social media has become a part of my daily routine
Smattitude4 I feel out of touch when I haven’t logged onto social media for a while
Smattitude5 I feel I am part of the social media community
Smattitude6 I would be sorry if social media websites shut down

Attitude toward brands
Smbrandatt1 I have a special bond with the brands that I like
Smbrandatt2 I consider my favorite brands to be part of myself
Smbrandatt3 I often feel a personal connection between my brands and me
Smbrandatt4 Part of me is defined by important brands in my life
Smbrandatt5 I feel as if I have a close personal connection with the brands I most prefer
Smbrandatt6 I can identify with important brands in my life
Smbrandatt7 There are links between the brands that I prefer and how I view myself
Smbrandatt8 My favorite brands are an important indication of who I am

Attitude toward social media advertising
Smadatt1 Interesting–Not Interesting
Smadatt2 Enjoyable–Not Enjoyable
Smadatt3 Informative–Not Informative
Smadatt4 Believable–Not Believable

Table AI.
Scale items used in the data collection for this project

(continued)
Social media intrusiveness

When I am on social media sites, advertising is…

<table>
<thead>
<tr>
<th>Smadintrusion1</th>
<th>Distracting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smadintrusion2</td>
<td>Disturbing</td>
</tr>
<tr>
<td>Smadintrusion3</td>
<td>Forced</td>
</tr>
<tr>
<td>Smadintrusion4</td>
<td>Interfering</td>
</tr>
<tr>
<td>Smadintrusion5</td>
<td>Intrusive</td>
</tr>
<tr>
<td>Smadintrusion6</td>
<td>Invasive</td>
</tr>
<tr>
<td>Smadintrusion7</td>
<td>Obtrusive</td>
</tr>
</tbody>
</table>

Corresponding author
Laura Frances Bright can be contacted at: l.bright@tcu.edu

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
In Investigating microblogging addiction tendency through the lens of uses and gratifications theory

Qian Li
Renmin University of China, Beijing, China

Xunhua Guo
School of Economics and Management, Tsinghua University, Beijing, China

Xue Bai
Department of Marketing and MIS, Fox School of Business and Management, Temple University, Philadelphia, Pennsylvania, USA, and

Wei Xu
Renmin University of China, Beijing, China

Abstract

Purpose – Considering the popularity and addictive attributes of microblogging, the purpose of this paper is to explore the key drivers of the microblogging addiction tendency, and to investigate the causal relationship between microblogging usage and addiction tendency through the lens of the uses and gratifications (U&G) theory.

Design/methodology/approach – By extending the U&G theory to accommodate the negative consequences of gratification, a research model that explains the relationships among microblogging use, gratification and addiction tendency was developed and empirically examined based on the data collected from 520 microblogging users in China.

Findings – The results showed that different types of microblogging use lead to different categories of gratification to different extents, while different categories of gratification play different roles in determining the level of addiction tendency. Specifically, the effect of content gratification on addiction is marginal, while social gratification has significant effects on all dimensions of addiction tendency.

Originality/value – The present study has both theoretical and practical implications. From a theoretical perspective, unlike many previous studies applied the U&G theory to explore the positive outcomes of media uses, this paper extends the U&G by including addiction tendency as a negative psychological outcome of U&G, resulting a research framework (use-gratification-addiction framework). Meanwhile, this paper contributes to the extending literature by examining the constructs of U&G at a granular level and investigated the causal relationship between “uses” and “gratifications.”

Keywords Uses and gratifications, Diminished impulse control, Microblogging addiction tendency, Microblogging gratifications, Microblogging uses, Social comfort

Paper type Research paper

1. Introduction

Studies of the factors driving the adoption and continued use of information technology have extensively investigated the positive outcomes resulting from its use, whereas the negative effects have not yet been sufficiently addressed (Turel et al., 2011). While the existing...
literature has indicated that intense use of information technology can have harmful consequences, such as reduced work performance, compromised social life and even severe mental problems (Jiang, 2014; Yang, Liu and Wei, 2016), the scope and formation of the negative effects have not been well conceptualized. Various terms have been adopted to designate the phenomena above, including technology addiction, problematic (or pathological) internet use, internet addiction disorder, compulsive use, computer addiction and technology dependency (Charlton and Danforth, 2007; Lu and Wang, 2008; Yang, Liu and Wei, 2016).

Even in the medical field, agreement has not been reached regarding a uniform term for such a phenomenon (Turel et al., 2011). To avoid the potential confusions with medical terms, instead of addiction, this paper focuses on addiction tendency, which refers to a user's psychological tendency toward maladaptive dependency, manifested through obsessive technology use at the expense of other important activities.

Technology addiction has been examined in contexts including online games (Charlton and Danforth, 2007; Lu and Wang, 2008; Xu et al., 2012), the internet (Iskender and Akin, 2011; Jia et al., 2007; Jiang, 2014), social networks (Thadani and Cheung, 2011), mobile e-mail (Turel et al., 2008), instant messaging (Huang and Leung, 2009) and online auctions (Turel et al., 2011). As a major form of social media service, microblogging platforms, such as Twitter, which allow users to publish short content via various means, has experienced skyrocketing growth within the last several years and has become one of the most popular internet services (Liu et al., 2011). In comparison with other existing internet or social networking services, microblogging has unique features, including its limit on post length, its support for mobile clients and its use of open publishing on the Web (Oulasvirta et al., 2010). These characteristics potentially increase the addictiveness of microblogging. First, regarding the length limit, microblogging aids users in obtaining highly fragmented and updated information. The more often that users employ microblogging, the more knowledgeable that they might feel (Li et al., 2017). Second, support for mobile clients indicates that a user's expression can receive rapid feedback from others anywhere and at any time, intensifying the user's enthusiasm into deep devotion (Li et al., 2012). Third, as a type of a user-based website, microblogging also provides social content, such as a social network, which encourages users to become highly involved in social connections (Thadani and Cheung, 2011). Previous studies also showed that high involvement in social networks may lead to addiction dependency (Yang, Wang, Lu, 2016; Xu and Tan, 2012). Additionally, microblogging allows users to build a self-selected network of people whom they follow without the network members' permission. The uncertainties of what will occur or what others will say on microblogging cultivate excitement and curiosity in users, resulting in intense use. Therefore, we choose microblogging as a context for the study of technology addiction.

To provide effective means to prevent or reduce addiction, previous studies have identified some antecedents of technology addiction, such as personal traits (Huang and Leung, 2009; Iskender and Akin, 2011; Jia et al., 2007), habits and attitudes (Turel and Serenko, 2011; Xu and Tan, 2012). However, the causal relationship between uses and addiction tendency has not yet been systematically addressed. Since many studies have suggested that addiction is a consequence of excessive use (Charlton and Danforth, 2007; Turel et al., 2008; Young, 1998), it is worthwhile to investigate the behavioral and cognitive intermediaries between use and the psychological state of addiction, which forms the focus of our study.

This paper aims to postulate a research model to understand the key drivers of microblogging addiction tendency. Our work draws upon a widely cited framework, the uses and gratifications (U&G) theory. The U&G theory framework fits well with the motivational perspective (Stafford and Gillenson, 2004) due to its inherent capability to explain the relationship between psychological status and use in accordance with addiction. We extend the U&G framework to include addiction tendency as a negative psychological outcome of U&G. Our analysis shows that the impacts of microblogging use on addiction
tendency are mediated by the gratification that a user gains from the use. Furthermore, we analyze the specific links from use types to gratification categories and from gratification categories to microblogging addiction tendency dimensions, resulting in a specialized model to explain the relationships among microblogging use, gratification and addiction tendency.

Our model is empirically tested using a data set including 520 microblogging users who used microblogging in diverse frequencies. The results show that different types of microblogging use lead to different categories of gratification to different extents. In particular, according to our study, use for the purposes of sharing and obtaining information has significant effects on both content and social gratification. The impact of entertainment use is strong and significant on both process and social gratification. Alternately, users gain the least gratification through microblogging use for the purpose of social communication, while different categories of gratification play different roles in determining the levels of addiction tendency. Specifically, the effect of content gratification on addiction tendency is marginal, while social gratification has significant effects on all dimensions of addiction tendency.

Our work contributes to the literature on U&G and to the technology addiction research in the following dimensions. First, our work represents a first attempt to disentangle the “use” and “gratification” components using the U&G framework. Few studies in U&G have examined constructs of U&G at a granular level and examined the causal relationship between “use” and “gratification.” Using microblogging as a showcase, we further conceptually define and demonstrate that various types of use lead to different categories and levels of gratification. Second, we postulate and validate a research framework that consolidates the causal relationships among use, gratification and addiction tendency. To the best of our knowledge, no research thus far has investigated the causal links from use through gratification to addiction tendency. Our research constitutes a first attempt in use, gratification and technology addiction research. Our findings yield important practical implications regarding whether and how addiction tendency can be reduced or properly channeled in personal and organizational environments.

2. Literature review

2.1 Technology addiction research

Technology addiction defines the degree of users’ psychological dependency through technology use (Turel et al., 2011). The negative effects of technology addiction on one's personal life (e.g. disturbance of family relationships, physical risk, pathological problems and isolation) and professional life (e.g. academic probation, work distraction, lessened productivity) have been widely recognized and analyzed (Jiang, 2014; Lu and Wang, 2008; Yang, Liu and Wei, 2016).

A significant implication for the study of technology addiction is to present solutions to reduce and then limit addictive use to a healthy level, indicating another research realm: the antecedents of addiction tendency. Because addiction tendency is a person’s psychological status with regard to technology, we group the antecedents into two categories (person-related and technology use-related) (see Table I). Noting the different addiction tendency levels among various users, personal demographics and traits were assumed to influence one’s psychological status during use (Jia et al., 2007). With regard to person-related antecedents, the effects of personal characteristics, such as self-compassion and shyness, self-concealment and alienation, on technology addiction tendency have been discussed (Huang and Leung, 2009; Iskender and Akin, 2011; Magsamen-Conrad et al., 2014). Previous works have focused on explaining the types of personal traits that have greater potential for addiction tendency. However, the impact of personal traits on addiction tendency cannot be achieved without technology use.

In terms of use-related factors, some research has examined the relationships between addiction and adoption, as well as continuance, attributed to crucial constructs such as habit and attitude (Xu and Tan, 2012). The level of use, represented by use rate and variety, is also
regarded as an antecedent of addiction tendency (Theotokis and Doukidis, 2009). Additionally, the user’s motivation, along with prevention factors regarding game addiction, have been discussed (Xu and Yuan, 2008; Xu et al., 2012). According to Whang et al. (2003), addicted, possibly addicted, and non-addicted users exhibit differences in internet use time length and use frequencies due to diverse motivations. Turel (2015) analyzed the use in five dimensions and validated their positive effects on addiction. These studies only employed use-related factors in the analysis of addiction tendency. However, to be able to explain how addiction tendency develops through use, concepts describing users’ psychological and cognitive states should be considered as well.

In summary, technology addiction tendency deserves more attention from MIS scholars, primarily in terms of the following aspects. First, in addition to the traditional contexts, additional research contexts must be explored. Alongside the development of technology, novel services have emerged with uniqueness, bringing new addictive attractiveness to users and thus challenges for future studies. Second, the role of use in affecting the addiction tendency level should be further explored. In previous studies, on the one hand, IT/IS use contributed to desirable positive outcomes, as widely discussed in the continuous usage field. On the other hand, in the research field of IT addiction tendency, use has been identified as a positive factor that increases users’ addiction tendency levels (Theotokis and Doukidis, 2009; Whang et al., 2003; Xu et al., 2012). These conflicts create the need for additional study efforts to identify the type of use that is crucial in affecting the addiction tendency. Third, understanding the mechanism that links IT/IS use and addiction tendency is crucial for addiction prevention. Although previous studies verified the influence of use on addiction tendency, they exerted little effort to explain the detailed mechanism.

2.2 U&G theory

The U&G theory, which is rooted in mass media, provides a user-centered perspective to explain that the user’s selection of certain media depends on the gratifications gained or, in other words, how well his/her needs are met (Rubin, 1979; Stafford and Gillenson, 2004). It was first developed in the 1940s for the study of radio audience satisfaction, followed by the study of various traditional media, such as newspapers, television, VCRs and electronic bulletin boards, over the last century (Cho et al., 2003). The internet resulted in an upsurge in U&G research because of the match between the internet’s strength with regard to interactivity and U&G’s emphasis on users’ activeness (Kaye and Johnson, 2004). A growing number of studies used this theory to explain individuals’ use of the internet, blogging, online games and virtual communities, as well as social networks (Liu et al., 2017). In particular, previous scholars also argued that this theory is suitable for the study of microblogging (Liu et al., 2010).

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Description</th>
<th>Context</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person-related</td>
<td>Characters</td>
<td>Demographics, personal traits</td>
<td>Instant message, internet</td>
<td>Jia et al. (2007), Huang and Leung (2009), Iskender and Akin (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Online games</td>
<td>Xu et al. (2012), Xu and Yuan (2008)</td>
</tr>
<tr>
<td></td>
<td>Motivation</td>
<td>Advancement, relationship, escapism</td>
<td>Online games</td>
<td>Turel and Serenko (2011), Xu and Tan (2012)</td>
</tr>
<tr>
<td>Use-related</td>
<td>Adoption/continuance</td>
<td>Attitude in TAM</td>
<td>SNS</td>
<td>Turel (2015)</td>
</tr>
<tr>
<td>Use level</td>
<td>Use level</td>
<td>Use rate and variety, use frequency, use duration, comprehensiveness</td>
<td>SNS</td>
<td></td>
</tr>
</tbody>
</table>

Table I. Antecedents of technology addiction tendency
U&G holds that users employ certain media to gratify one or more of their needs (Chen, 2011). Along with the study of different media contexts, abundant typologies of U&G have been elicited, including entertainment, social interaction, escape, convenience and information (Lim and Ting, 2012; Liu et al., 2010). In addition to detailed discussions, some researchers have analyzed U&G in categories. In earlier studies of U&G, which primarily focused on traditional media, two main categories were conceptualized: content gratification, which indicates that users employ media for the content carried by a medium (e.g. information, entertainment); and process gratification, which indicates that users gain through the use experience itself (e.g. escape, surfing) (Cutler and Danowski, 1980; Stafford and Stafford, 1996). Considering the novelties of the internet, social gratification was suggested to generalize satisfaction with regard to media’s social environment, which cannot be expressed well using content or process gratification (Stafford and Gillenson, 2004). Moreover, technology gratification was introduced by Liu et al. (2010) to describe a suitable and convenient technology environment.

3. Research model and hypotheses
3.1 Use-gratification-addiction (U-G-A) framework
The U&G theory emphasizes the correspondence between intended uses and resulting gratifications (Stafford and Gillenson, 2004). Weibull (1985) noted that people return to a medium that they find gratifies their needs, indicating that repeated media use induces constant gratification. This constant feeling leads to psychological reinforcement and then eventually results in dependence (Hanson et al., 2008). In other words, users become dependent on something for the satisfaction and psychological reinforcement that it provides. When technology addiction tendency is interpreted as a negative psychological dependency caused by the use of the technology, in the lens of the U&G theory, it is reasonable to expect that gratification plays a mediating role between use and addiction tendency. Therefore, users with high gratification levels have a greater possibility of becoming addicted. Based on the analysis, a framework called U-G-A, presented in Figure 1, is created.

Considering the phases of use, the framework above can be illustrated using a multi-stage expression. U&G emphasizes that users are goal directed and that media is used to gratify needs, such as information, social connection, etc. (Katz et al., 1973; Wu et al., 2010). Therefore, the types of use are applied in the framework to represent the different uses, corresponding to the first stage. With respect to U&G, gratification is defined as some aspect of satisfaction related to active use (Stafford and Gillenson, 2004), namely, gratification gained during or after use (Cho et al., 2003), which could be conceptualized as the second stage. Weibull (1985) noted that people return to a medium that they find gratifies their needs, indicating that repeated media use induces constant gratification. This constant feeling leads to psychological reinforcement and then eventually to dependence (Hanson et al., 2008). In other words, users become dependent on something for the satisfaction and psychological reinforcement that it provides (Meyer and Dibbern, 2010). Consistent with the definition of technology addiction tendency as a psychological

![Figure 1. The U-G-A framework](image-url)
dependency, the third stage in our framework offers the proposition that users with high gratification have a greater possibility of becoming addicted, with negative consequences for their other responsibilities.

Applying U-G-A framework in a microblogging context, microblogging use, gratification and addiction tendency are discussed in detail in the following portion of the paper.

3.2 Microblogging uses
Technology use can be categorized according to specific activities (i.e. various operations) or use types (i.e. the different needs being met) (Li et al., 2013). With regard to addiction tendency, most studies have tended to focus on use types rather than activities. For instance, the types of use, such as advancement, mechanics, relationships and escapism, were shown to have effects on online game addiction (Xu et al., 2012). In this sense, it can be inferred that microblogging addiction tendency is fostered through different types of use by the participants, such as information sharing and self-expression, instead of posting, reading or other specific activities. Several studies have been conducted with general analysis of microblogging use (Java et al., 2007; Zhao and Rosson, 2009), in which detailed types of use, such as chatting and news reporting, have been studied. In this paper, we identify three types of microblogging use – social communication, information and entertainment – as detailed in the following discussions.

3.2.1 Social communication. The use of social communication, connection and interaction appears to be prominent and significant in microblogging. Through the use of microblogging, people conduct ordinary activities and have experiences visible to others (Oulasvirta et al., 2010), maintain ongoing relationships (Dunlap and Lowenthal, 2009), present oneself, contact friends and family (Sammer and Back, 2011), conduct informal communication (Meyer and Dibbern, 2010) and become connected to certain interest networks (Kim et al., 2010). Therefore, we identify “social communication” as a microblogging use type to conceptualize users’ exchanges of personal status and social interactions at the relationship and community levels.

3.2.2 Information. Regarding microblogging as a type of social media, “information” is presented as another type of use to represent information sharing and consumption among users. Information being shared and consumed includes news, resources and knowledge. For example, according to research by Oulasvirta et al. (2010), users share opinions, information and links to resources via microblogging. Providing users with short messages that support awareness, microblogging is used to share tacit and explicit knowledge (Meyer and Dibbern, 2010). Moreover, people share ideas and disseminate information via microblogging (Honeycutt and Herring, 2009) and engage in quick reflections and discussions (Ebner and Schiefner, 2008).

3.2.3 Entertainment. In addition to social communication and information sharing and consumption, other types of use have been discussed in academic research, such as the release of emotional stress (Zhao and Rosson, 2009) and the passing of time (Liu et al., 2010), as well as in commercial reports, e.g. celebrity worship (iResearch, 2012). Many studies in the U&G realm have broadly identified such entertainment uses as an important factor affecting the use of a particular medium (Liu et al., 2010; Luo et al., 2006). In this paper, we define the third type of microblogging use, “entertainment,” to represent these pleasurable and relaxing uses, usually without a meaningful purpose, such as the passing of time, leisure, celebrity worship, etc., which are very popular in microblogging use according to a survey conducted in China (iResearch, 2012).

3.3 Microblogging gratifications
Most existing U&G research has not distinguished between U&G, with gratification typically defined as aspects of satisfaction related to media use (Stafford and Gillenson, 2004). In this
paper, we define gratification as the satisfaction gained through microblogging use, emphasizing the causal relationship between U&G. Content, process and social gratifications are representative gratification categories (Stafford and Gillenson, 2004) that are widely accepted in U&G research (Chen, 2011; Liu et al., 2010).

3.3.1 Content gratification. Content gratification is conceptualized when people obtain satisfaction from the content carried by a medium (Cutler and Danowski, 1980; Stafford and Stafford, 1996). As a new Web 2.0 platform, microblogging provides abundant content, such as words, images and resource links, with a great number of participants distributed across different continents (Java et al., 2007; Meyer and Dibbern, 2010). Users derive value from the content according to their own interests, inducing content gratification.

3.3.2 Process gratification. Process gratification has been suggested to conceptualize the user’s satisfaction with the pleasurable and flexible experience of media use (Cutler and Danowski, 1980; Stafford and Stafford, 1996). In the microblogging context, without permission from others, every user owns a special self-selected network created completely as he/she likes, leading to enjoyment and curiosity. Furthermore, support for mobile devices provides a high degree of convenience, prompting users to obtain gratification from the use process.

3.3.3 Social gratification. Social gratification has been used to conceptualize the user’s satisfaction with the social environment (Stafford and Gillenson, 2004). People become satisfied by presenting themselves and forming relationships with other participants via microblogging (Chen, 2011). In particular, without other participants’ permission, users can build self-selected networks on microblogging platforms. Users gain the gratification of belonging to certain groups, such as groups interested in the solar system, and they obtain respect from others by shaping themselves as specialists regarding a particular problem. In other words, users attain gratification within the social environment provided by microblogging.

3.4 Microblogging U&G

Media with high interactivity is likely to satisfy users. Microblogging provides utilities with which users can form social connections at personal, relational and community levels. The gratification of users’ needs to connect with others is fostered on Twitter, a similar social networking site (SNS) to microblogging (Chen, 2011). Magsamen-Conrad et al.’s (2014) study provided evidence that social communication through a mediated channel helps some individuals to foster beneficial interpersonal relationships. Additionally, the real-time feedback from microblogging increases users’ social interactions (Liu et al., 2011). By satisfying users’ need for social interaction, social communication use produces social gratification (Liu et al., 2010). Hereby:

H1. The use of microblogging for social communication has a positive effect on the level of social gratification.

Providing users with short messages that support awareness, microblogging is used to share tacit and explicit knowledge (Meyer and Dibbern, 2010). Moreover, Twitter can be used to share ideas and disseminate information (Honeycutt and Herring, 2009), as well as news and URL resources (Java et al., 2007). Content gratification increases with users’ satisfaction with the content carried by a medium (Stafford and Stafford, 1996). As the most popular and fastest-growing social medium, microblogging provides extremely abundant information, satisfying participants’ need for content. On the other side, information sharing in microblogging is different from other media. Since information is shared through a network built according to one’s social connections or interests, sharing increases users’ social gratification. In particular, using microblogging for information sharing, users, such as members of a science interest group, also interact with one another and might feel
satisfied with regard to social connections, as well as membership. At the same time, through use for the purpose of information sharing and consumption, the user might feel better informed with updated news from professional areas or social circles in which he/she is interested, therefore being gratified by the feelings that he/she is timely and closely involved in the development frontiers of the social groups to which he/she belong. Moreover, microblogging participants can easily share their knowledge, ideas and resources worldwide and thus gain the opportunity to become respected specialists in a particular field (Li et al., 2012), increasing users’ satisfaction with social influence. Hence:

\( H2a. \) The use of microblogging for information has a positive effect on the level of social gratification.

\( H2b. \) The use of microblogging for information has a positive effect on the level of content gratification.

Sometimes, participants use microblogging simply for leisure purposes, such as the release of emotional stress (Zhao and Rosson, 2009) and the passing of time (Liu et al., 2010), which can be characterized as entertainment motivations. As noted in a previous study on media use, such activities are associated with process gratification since they focus on deriving pleasurable outcomes directly from the act of media consumption (Song et al., 2004). Their needs are satisfied through experience in the process of use, which is a typical form of process gratification (Stafford et al., 2004). In Young’s (1998) research, it was also found that process gratification is more highly associated with entertainment-seeking activities, such as online chatrooms and games, rather than information-oriented activities, such as news browsing. At the same time, the one-directional confirmation mechanism used in microblogging allows participants to freely follow certain users as entertainment resources. Consequently, users access social interaction with others with similar entertainment resources, indicating social gratification. For instance, by following celebrities whom they admire, users feel connected to other fans. In this sense, the entertainment-oriented activities of microblogging are associated with pursuing the feelings that one has interactions with the celebrities that he/she worships, as well as with others who admire the same celebrities. This argument is supported by the findings from a previous study emphasizing that entertainment activities such as celebrity worship take on a social characteristic that involves not only individualistic behaviors, such as watching and reading about celebrities, but also an emotional exchange between the fans and the celebrity (McCutcheon et al., 2002). Therefore, we have the following hypotheses regarding entertainment:

\( H3a. \) The use of microblogging for entertainment has a positive effect on the level of social gratification.

\( H3b. \) The use of microblogging for entertainment has a positive effect on the level of process gratification.

3.5 Microblogging addiction tendency

In line with the definition of addiction tendency, microblogging addiction tendency is defined as a user’s psychological dependency on microblogging through use at the expense of other duties. In our model, microblogging addiction tendency is treated as a four-dimensional construct proposed by Davis et al. (2002), namely, diminished impulse control, loneliness/depression, social comfort and distraction. The dimension of diminished impulse control involves obsessive cognitions. In this paper, it represents users’ inability to control their microblogging use rationally. Loneliness/depression involves feelings of worthlessness and depression. In our context, it illustrates users’ negative emotions related to banning or limited microblogging use. Social comfort involves the feelings of safety and
security in a social network. In our case, it refers to users’ feelings of comfort and carefreeness regarding microblogging, comparing with their feelings in reality. Distraction involves avoidance through use. In our context, it emphasizes the users’ dependency on microblogging for procrastination.

We choose the four-dimensional construct proposed by Davis et al. (2002) for the following reasons. First, as mentioned above, the study of addiction has not yet arrived at uniformity in terms. Although the phenomenon is conceptualized as problematic internet use, instead of addiction tendency, in Davis et al. (2002)’s work, the four measurement dimensions cover users’ psychological dependency and negative consequences, consistent with our definition. Meanwhile, we regard microblogging addiction tendency as a continuous concept with multiple dimensions, which can be specialized through the construct of Davis et al. (2002).

Some scholars have classified microblogging platforms (such as Twitter) as SNS in addiction research (Thadani and Cheung, 2011; Theotokis and Doukidis, 2009; Xu and Tan, 2012). However, in our opinion, microblogging has characteristics that cannot be represented by SNS. The effects of social interaction, connection and communication among users are regarded as the outstanding characteristics of SNS addiction (Xu and Tan, 2012), and they can also be verified in the microblogging context. Nevertheless, as not only a social network but also as a medium, microblogging has a significant information diffusion component, which has been discussed above. Additionally, in contrast to SNS, microblogging attracts users from a wider range of ages and industries and has more effects on society and the economy considering special users, such as the government and enterprises. Finally, unlike social networks, in which requests for connection to others require permission, the “follow” action of microblogs is a one-directional confirmation mechanism, providing users with totally novel experiences, such as celebrity worship by “following” celebrities, which can hardly be realized in the real world or with other internet services.

3.6 Microblogging gratifications and addiction dependency
3.6.1 Social gratifications and addiction dependency. The existing literature has revealed that the fulfillment of users’ needs for social relationships increases online game addiction levels (Xu et al., 2012). Furthermore, social influence has indirect but positive influences on social information system addiction tendency (Theotokis and Doukidis, 2009). In general, the sense of community is positively related to the tendencies of addiction (Song et al., 2004). Similarly, Yang, Liu and Wei’s (2016) analysis showed that social interaction positively affects mobile SNS addiction tendency.

As mentioned above, microblogging addiction tendency is analyzed in four dimensions in this paper. Correspondingly, the effects of social gratification on diverse addiction tendency dimensions are studied. First, a high level of social gratification indicates a user’s high degree of satisfaction with ongoing interactions. Supported by mobile clients, users can receive social feedback via microblogging anytime and anywhere (Liu et al., 2011), increasing the difficulty of voluntarily controlling use. In other words, accumulated social gratification gained through staying online all the time would tend to cause the users to keep thinking about their experiences with microblogging and to become unable to rationally cut down on their use, shaping a typical characteristic of “diminished impulse control” (Davis et al., 2002). Second, microblogging participants maintain their connections to family, friends and other users within their self-selected networks; even those with certain traits, such as shyness, feel much less inhibited in social interactions online than they do offline (Huang and Leung, 2009). Similarly, Twitter provides a high degree of interactivity (Liu et al., 2010), i.e. cutting use of Twitter means interrupting users’ interactivity with others.
According to Chen (2011), users stick with Twitter to consistently satisfy the gratification of being connected with others. Consequently, limitation of microblogging use will affect the user’s connections to others, reducing the level of social gratification, and then one’s loneliness/depression will increase, typically reflecting the characteristics of loneliness/depression (Davis et al., 2002). Third, when a user gains a high level of social gratification via microblogging, he/she may feel more and more comfortable with the virtual social environment enabled through the use. As mentioned above, considering the self-selected networks in microblogging, membership and influence are identified as two key components of social gratification. The intense feeling of belonging to certain groups (i.e. membership) implies a high level of emotional safety. Meanwhile, one’s influence represents his/her respect among users. Conceptually, the feelings of comfort, respect and safety online are key elements of the construct of “social comfort” (Davis et al., 2002). Finally, as discussed above, social gratification represents a user’s evaluation of the connections within a personally selected network, which usually does not promote one’s work or study duties and, contrarily, increases one’s intentions to ignore his/her responsibilities. As microblogging allows users to stay online all the time, users are in fact tightly bound to the virtual social environment through the mobile devices that they always carry (Liu et al., 2011). In this sense, microblogging keeps attracting the user to enter the social environment whenever possible, creating a strong force toward distraction, which indicates that one becomes increasingly unable to concentrate on his/her own responsibilities (Davis et al., 2002). Hence, we conclude from the discussions above the following four hypotheses regarding social gratification:

- **H4a.** Social gratification is positively related to diminished impulse control.
- **H4b.** Social gratification is positively related to loneliness/depression when use is restricted.
- **H4c.** Social gratification is positively related to social comfort.
- **H4d.** Social gratification is positively related to distraction.

### 3.6.2 Content gratifications and addiction dependency

It was proposed in a previous study that content gratification was unrelated to addiction tendencies (Song et al., 2004). Partially along with this opinion, we argue that content gratification is not related to loneliness/depression or distraction based on the following analysis. Content gratification increases with content quality and has a weak relationship with personal emotions, i.e. loneliness/depression. At the same time, users at a high level of content gratification can sometimes obtain support for work-related problem solving, whereby microblogging promotes one’s concentration on his/her duties rather than distracting from these responsibilities.

Alternately, considering the characteristics of microblogging, we argue that the content gratification is related to the other two dimensions of addiction tendency: diminished impulse control and social comfort. In detail, users attain content gratification according to their own preferences and interests without time or geographical constraints; consequently, their capabilities for impulse control might unconsciously be decreased during usage. Furthermore, a high degree of content gratification indicates that users are engaged in content sharing and acquisition on the Web, offering them the feeling of being knowledgeable and bringing them respect and safety, which can rarely be assured in real life. In other words, content gratification can lead to social comfort on microblogging. Thus:

- **H5a.** Content gratification is positively related to diminished impulse control.
- **H5b.** Content gratification is positively related to social comfort.
3.6.3 Process gratification and addiction dependency. Process gratification represents users’ evaluations of experiences, which are usually pleasurable and enjoyable. Previous studies have demonstrated that, apart from any consequences, the feeling of enjoyment of the use process positively influences Web shoppers with regard to use attitudes and causes them to return to a site (Cyr et al., 2007). Similarly, enjoyment significantly affects mobile SNS addiction (Yang, Wang, Lu, 2016). Another study has also empirically validated that factors representing process gratification are significantly related to addiction tendency in the case of internet use (Song et al., 2004).

In the context of microblogging, users gain process gratification more easily and frequently with support from mobile handheld devices, indicating difficulty in use control. Additionally, microblogging users obtain process gratification within their self-selected networks in accordance with their interests and favorites, which is not possible with other services. If the use is limited, meaning that exploration of one’s own network is interrupted, loneliness or depression is induced. Furthermore, users with a high degree of process gratification indicate their positive evaluations of the experience of microblogging use. Particularly through the assistance of mobile clients, users’ attentions are shifted from their jobs or other duties through their continuous use of microblogging. Therefore:

- **H6a.** Process gratification is positively related to diminished impulse control.
- **H6b.** Process gratification is positively related to loneliness/depression when the use is restricted.
- **H6c.** Process gratification is positively related to distraction.

3.7 Research model
According to the discussions above, a research model is postulated based on the general U-G-A framework to determine the specific relationships among each type of use, shown in Figure 2.

4. Methodology
4.1 Measurement
To examine the proposed hypotheses, the measurements are developed based on the definitions of constructs regarding the context of microblogging.

4.1.1 Microblogging uses measurement. Users engage in communication and self-disclosure on microblogging (Oulasvirta et al., 2010; Meyer and Dibbern, 2010); thus, the use of “social communication” is formatively represented by two related constructs used
in previous studies, i.e. self-expression and social interaction (Liu et al., 2010). Items are adjusted to fit our research context.

The use type of “information” is used to encapsulate the exchange of news, knowledge, etc., on microblogging, and the formative measurements of this motivation should consider both sharing and consumption. Information sharing has been measured in many studies (Constant et al., 1994; Kolekofski and Heminger, 2003). However, the items have primarily emphasized willingness to and attitudes toward information sharing and are thus unsuitable for representing the use types. In this paper, the measurements of information sharing are adapted from Jarvenpaa and Staples’ (2000) work, which focused on the detailed content of information. Conversely, information acquisition has primarily been discussed in studies related to learning (Vandenbosch and Higgins, 1996), and measurements are adjusted in this paper to be consistent with microblogging use.

The third use type, “entertainment,” has been widely accepted as a major component in media use with well-developed measurement items (Liu et al., 2010; Luo et al., 2006). Considering the uniqueness of microblogging use, entertainment is measured using two formative indicators in this paper, namely, leisure and celebrity worship.

Gratification measurement. Within the U&G research realm, scholars have developed measurement items for content and process, as well as social gratifications (e.g. Liu et al., 2010), in which the U&G were not clearly distinguished, resulting in difficulties in adopting them directly in our study. In our opinion, content gratification is determined by abundant information on microblogging. As a source of knowledge, resources, opinions, feedback and discussions, microblogging provides content gratification for users in the form of feelings that he/she is knowledgeable or experienced with work, study or other duties, which can be reflectively measured using a modification of similar constructs, such as perceived relative advantage (Moore and Benbasat, 1991).

Process gratification is obtained through pleasurable and flexible microblogging use experience, particularly in terms of the feelings of enjoyment, curiosity and convenience associated with the use process. Enjoyment was widely discussed in previous studies as a component of gratification in almost every media context (Lim and Ting, 2012; Liu et al., 2010). From the perspective of addiction tendency, enjoyment plays a more important role, whereas other feelings, such as curiosity and convenience, have received comparatively less attention in causing users to commence intense use. Therefore, in this paper, process gratification is reflectively measured based primarily on the work of perceived enjoyment (Cyr et al., 2007).

Social gratification has usually been measured by social interaction or social presence in previous studies (Liu et al., 2011; Wu et al., 2010). In addition to social presence, microblogging platforms provide self-selected networks, allowing users to find and “follow” others who have similar interests, opinions, backgrounds, etc., with social gratification being represented primarily by feelings of influence and membership (Koh and Kim, 2003). Consequently, social gratification is formatively measured using social presence, perceived membership and perceived influence.

Microblogging addiction tendency measurement. Microblogging addiction tendency is measured in terms of four discrete constructs: diminished impulse control, loneliness/depression, distraction and social comfort. All of these items are modified to reflect the perspectives of psychological dependencies in accordance with the definition of addiction tendency. All of the measurement items for the constructs in the research model are listed in Table A1.

Sample
To validate the research model, a survey was conducted in China in November 2012. After excluding from the sample 35 participants who were not microblogging users, who did not
finish the questionnaires or whose responses to all of the questions were the same, a total of
520 valid responses were obtained with the survey. Among the responses, 38 percent were
from students at a university in China, as well as their families, friends and other social
connections approached using personalized e-mail, whereas the majority (62 percent) of the
respondents were invited by a professional company to visit a website containing our
questionnaire. In previous studies, a large, at times even 100 percent, proportion of
respondents consisted of university students (Jia et al., 2007; Thadani and Cheung, 2011)
because young educated people have been viewed as more likely to be innovative and
involved in technology services. Considering that microblogging users vary widely
in terms of age and industry compared with users of other internet services, our sample is
more representative than purely student-based surveys. The entire questionnaire took
approximately 20 minutes to complete.

There was a larger proportion of male participants (57 percent) than female participants
(43 percent) in the survey, and all of them lived in China. According to a survey on the
distributions of the ages of microblogging users in China, the majority of microblogging
participants are approximately 20 years old (60 percent), but the substantial percentage of
users who are 30 years or older (33 percent) demonstrates the importance of older users as
well (data source: www.enfodesk.com). The age distribution of the 520 respondents in our
study indicated accordance with this commercial survey and represented the variety of
education and use history. Among the respondents, more than 80 percent of users use
microblogging every day, and nearly half of the sample spend more than 1 hour per day on
it, which is suitable for our study of addiction tendency. More detailed demographic
information is provided in Table II.

5. Data analysis and results
5.1 Validity and reliability
The internal consistency reliability of the study was assessed by computing Cronbach’s α,
as shown in Table III. Generally, a Cronbach’s α greater than 0.7 indicates acceptable
reliability (Gefen and Straub, 2005). In our study, the Cronbach’s α values ranged from
0.86 (for distraction) to 0.96 (for process gratification), indicating that the reliability is
acceptable. Table III also shows the values of the average variance extracted (AVE) used to

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category</th>
<th>Frequency (percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>18 or younger</td>
<td>30 (5.8%)</td>
</tr>
<tr>
<td></td>
<td>19–24</td>
<td>188 (36.2%)</td>
</tr>
<tr>
<td></td>
<td>25–30</td>
<td>146 (28.1%)</td>
</tr>
<tr>
<td></td>
<td>31–35</td>
<td>75 (14.4%)</td>
</tr>
<tr>
<td></td>
<td>36–40</td>
<td>37 (7.1%)</td>
</tr>
<tr>
<td></td>
<td>40 or older</td>
<td>44 (8.5%)</td>
</tr>
<tr>
<td>Highest level of education</td>
<td>Junior middle school</td>
<td>5 (1%)</td>
</tr>
<tr>
<td></td>
<td>High school</td>
<td>34 (6.5%)</td>
</tr>
<tr>
<td></td>
<td>Junior college or Pre-U</td>
<td>157 (30.2%)</td>
</tr>
<tr>
<td></td>
<td>Bachelor</td>
<td>285 (54.8%)</td>
</tr>
<tr>
<td></td>
<td>Master or above</td>
<td>39 (7.5%)</td>
</tr>
<tr>
<td>Microblogging usage</td>
<td>Less than 1 year</td>
<td>131 (25.2%)</td>
</tr>
<tr>
<td></td>
<td>1–3 years</td>
<td>350 (67.3%)</td>
</tr>
<tr>
<td></td>
<td>3–5 years</td>
<td>36 (6.9%)</td>
</tr>
<tr>
<td></td>
<td>5 years or more</td>
<td>3 (0.6%)</td>
</tr>
</tbody>
</table>

Table II. Sample distribution Note: n = 520
measure convergent validity. With all of the AVE values clearly greater than the required level of 0.50, the convergent validity of the research is satisfactory.

To assess the discriminant validity, the results of the confirmatory factor analysis are calculated and are shown in Table AII, indicating that the measurement items are essentially discriminated into the corresponding constructs in the model. Table IV provides the correlation matrix for the constructs.

5.2 Mediating effect of gratification
Before testing the proposed model, we first examine the mediating effect of gratification between use and addiction tendency. For this purpose, a two-step regression analysis is conducted. In the first step, we estimate the direct impacts of microblogging use on addiction tendency and gratification, including age, gender and education level as control variables. All three high-order constructs are computed as formative compositions of their sub-constructs. The results in Figure 3 show that microblogging use is positively related to both gratification and addiction tendency at a high significance level. Among the three control variables, only gender shows a significant impact on gratification such that men tend to have more gratification than women. Therefore, it is empirically evident that heavy microblogging uses induce high levels of gratification and addiction tendency.

Subsequently, a structural equation model is adopted to test the mediating effects of gratification. The results in Figure 4 illustrate that, after adding gratification as a mediating construct, the direct effect of microblogging use on addiction tendency is significantly reduced.
element, the direct impact of use on addiction tendency remains significant, but the coefficient drops by 59 percent to 0.304. More importantly, the impact of gratification on addiction tendency is validated at a high significance level. Compared with the results in Figure 3, the $R^2$ value of addiction tendency increases by 19 percent, indicating that the variances of addiction tendency are better explained. Based on the results, we argue that the overall addiction tendency level is highly correlated with the amount of use, whereas the correlation is significantly mediated by the gratification that a user gains from microblogging.

The results above strongly suggest that addiction tendency is a negative psychological outcome of use with the mediating effects of gratification, validating the U-G-A relationship framework. This framework indicates that gratification is obtained through use, and constant gratification leads to psychological dependency, in turn leading to addiction. Based on this foundation, we further test the specific relationships among types of use, categories of gratification and dimensions of addiction tendency.

5.3 Hypothesis testing
Table V summarizes the testing results of the research model. As indicated in the results, most of the hypotheses in the model are significantly supported, with the exception of the impact of content gratification on diminished impulse control. The variances of the additional dimensions are explained well, with all $R^2$ values greater than 0.40 (except that for process gratification), indicating a good fit overall. Therefore, it can be concluded that the postulated model largely captures the influence mechanism, from the microblogging use to the gratification gained through use and subsequently to the emergence of microblogging addiction tendency.

![Figure 3. Direct impacts of use on gratification and addiction](image)

![Figure 4. Mediating effect of gratification](image)
All the hypotheses regarding the relationships between U&G are supported at a significance level of $p < 0.001$ except $H1$ at 0.05. A possible explanation is that, in addition to the other two use types (information and entertainment) being able to lead to social gratification, other prevalent social media applications, such as Facebook and WeChat, also gratify users' social needs, causing the importance of the social use of microblogging to be relatively less.

Most of hypotheses on the relationships between gratifications and addiction tendency are also supported at high significance levels. The distraction effect ($H6c$) is moderately supported ($p < 0.05$), perhaps because process gratification generates pleasurable and flexible feelings and serves as temporary relief form, rather than avoidance of one's work. In addition, social and process gratifications explain 63 percent of the variance of loneliness/depression ($H4b$ and $H6b$), indicating that, if the use was interrupted, users who enjoy greater satisfaction through social interactions on microblogging or through the pleasurable blogging experience itself would have stronger feelings of loneliness and depression. The link between content gratification and diminished impulse control ($H5a$) is not supported. Considering that the $R^2$ of diminished impulse control is very high, it can be inferred that process and social gratifications are two dominant factors that lead to diminished impulse control, whereas the effect of content gratification is marginal compared with the other two gratification factors.

In summary, the postulated model is significantly supported by the empirical testing; the results indicate that the various types of microblogging use can lead to various types of gratification, which, in turn, facilitate the emergence of addiction tendency in various dimensions. Therefore, as gratification accumulates, addiction tendency emerges and grows, but the levels of addiction tendency within the four dimensions are not synchronous.

6. Discussion and conclusion

6.1 Discussion of the results

Considering the popularity and addictive attributes of microblogging, this study investigates the causal relationships among microblogging uses, gratifications and addiction tendency.
A research framework, i.e. the U-G-A framework, is introduced, which extends the U&G theory by focusing on the possible negative consequences of continued gratification. Based on the regression analysis, gratification is identified as having a mediating effect on the relationship between use and addiction tendency. Specifically, a model is introduced to describe the specific links from use types to gratification categories and from gratification categories to microblogging addiction tendency dimensions. According to our empirical tests, the use of information and entertainment provides users high evaluations on microblogging’s importance as a novel medium. However, the use of social connection plays a less important role in affecting gratification, indicating microblogging’s deficiency as a social network. Furthermore, our research validates the effects of social gratification on all addiction tendency dimensions, consistent with previous studies (Song et al., 2004; Xu et al., 2012). However, in contrast with the common intuition, process and content gratification have unbalanced influences on diverse addiction tendency dimensions, which could hardly be found if the addiction tendency were treated as a uni-dimensional concept.

6.2 Implications to research
Generally, based on theoretical research and a two-step regression analysis, our study validates the U-G-A framework as a consolidated theoretical tool for studying the negative outcomes of technology use. In this sense, our research contributes to the literature on the U&G theory by illustrating the casual relationships between U&G and by extending the U&G framework to accommodate the negative consequences of gratification, an aspect that has not been addressed in previous studies from a theoretical standpoint. Based on the U-G-A framework, our model conceptualizes the specific constructs in the context of microblogging and postulates causal links among the constructs. Our research is also among the first to investigate microblogging by identifying its unique characteristics within the conceptual domains of use types, gratification and addiction tendency, with the causal links among them postulated and verified.

6.3 Implications to practice
From a practical standpoint, the findings of the research shed light on the balanced use of microblogging. On the one hand, social gratification results in a greater possibility of addiction tendency along all of the dimensions. From a societal perspective, methods for improving social gratification in offline environments, such as investment in more sports and entertainment centers, should be strengthened. Although users can gain social gratification anytime and anywhere through microblogging, face-to-face communications can introduce users to real and richer experiences. On the other hand, content gratification significantly affects only one dimension of addiction tendency and can benefit users’ work performance. From educational or business organizations’ points of view, content gratification can be enhanced by applying microblogging as a learning and cooperating tool without worrying about inducing addiction tendency.

For service providers whose interests are attracting users, gaining competitive advantages and improving business performance, our research suggests that the services could be provided in a more socially responsible manner. Our findings show that content gratification has the least impact on addiction tendency, encouraging providers to enhance microblogging’s role as a platform of information and knowledge resource and to promote its healthy use at the organizational level. Moreover, the effects of uses on gratifications in our study imply potential improvements for enriching the user experience of social communication via microblogging to boost microblogging’s ability to attract individual users or business customers and to compete with other social services.
6.4 Limitations and future research
As an initial attempt to develop and validate the U-G-A theoretical framework for the study of technology addiction tendency, this work is prone to limitations that, in turn, offer opportunities for future research. Our current model embodies microblogging’s characteristics only by means of identifying corresponding dimensions and not by adding specific addictive features. The constructs in the model could be better measured based on the comparison among diverse instruments (Turel et al., 2011). Furthermore, future research could explore the relationships between microblogging addiction tendency and other types of addiction tendency (e.g. online games), as well as clinical comorbidities, aiming to verify whether the current conclusions still hold with other potential influencing factors included. In addition, the average time that users spent on microblogging every day could be collected to evaluate use frequency. However, in line with our definition of addiction tendency, use frequency must be further specialized according to the context. For instance, users’ distractions from their duties mainly rely on the time that they spend during work or study, which cannot be reflected by average use time. Correspondingly, in future studies, usage record data should be included to provide more accurate analyses on use frequency. Moreover, the measurements of addiction tendency applied in this paper mainly follow ideas from IS research. Considering the possible severe consequences of addiction tendency, the clinical comorbidities should be included among the measurements in future studies. Finally, data available through public sources are limited; therefore, we relied on a self-reported survey for measurements, which has certain limitations. One of the ongoing research directions is an attempt to incorporate usage record data as a complement to surveys and to evaluate the quantity of self-reports based on the corresponding usage record.

References
Hanson, G., Venturelli, J.P. and Fleckenstein, A.E. (2008), Drugs and Society, 10th ed., Jones and Bartlett Publishers, Sudbury, MA.


Further reading


### Table AI. Constructs and measurement items

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Formative components</th>
<th>Measurement items</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Types of use</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social communication (SC)</td>
<td>Self-expression</td>
<td>SE1: through the use of MB, I show my experience and status</td>
<td>Liu et al. (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SE2: I, through the use of MB, I express my feelings and emotions</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SE3: through the use of MB, I indicate my attitudes and thoughts</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social interaction</td>
<td>SI1: through the use of MB, I communicate with my friends and family</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SI2: through the use of MB, I keep in touch with my friends</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SI3: through the use of MB, I learn the statuses of my friends</td>
<td></td>
</tr>
<tr>
<td>Information (I)</td>
<td>Information acquisition</td>
<td>IA1: through the use of MB, I keep up with timely news</td>
<td>Vandenbosch and Higgins (1996), Jarvenpaa and Staples (2000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IA2: through the use of MB, I acquire resources (URLs)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IA3: through the use of MB, I obtain knowledge</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IA4: through the use of MB, I learn of comments on hot events</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Information sharing</td>
<td>IS1: through the use of MB, I broadcast news</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS2: through the use of MB, I share resources (URLs)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS3: through the use of MB, I broadcast others’ comments</td>
<td></td>
</tr>
<tr>
<td>Entertainment (ENT)</td>
<td>Leisure</td>
<td>LE1: through the use of MB, I pass time</td>
<td>Liu et al. (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LE2: through the use of MB, I escape</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LE3: through the use of MB, I chase for fun</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LE4: I use MB when I have nothing better to do</td>
<td></td>
</tr>
<tr>
<td>Celebrity worship</td>
<td>CW1: through the use of MB, I access celebrity news</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CW2: through the use of MB, I learn the statuses of celebrities</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CW3: through the use of MB, I learn of other fans’ love of celebrities</td>
<td></td>
</tr>
<tr>
<td>Gratification</td>
<td>Content gratification (C)</td>
<td>C1: the use of MB is advantageous for my work</td>
<td>Moore and Benbasat (1991)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C2: the use of MB makes my work more efficient</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C3: the use of MB improves the quality of the work I do</td>
<td></td>
</tr>
<tr>
<td>Process gratification (P)</td>
<td>P1: MB use is interesting</td>
<td>Cyr et al. (2007)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P2: MB use is entertaining</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P3: MB use is enjoyable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social gratification (S)</td>
<td>Social presence</td>
<td>SP1: there is a sense of human contact on MB</td>
<td>Liu et al. (2011), Koh and Kim (2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SP2: there is a sense of sociability on MB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SP3: there is a sense of human warm on MB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SP4: there is a sense of human sensitivity on MB</td>
<td></td>
</tr>
<tr>
<td>Perceived membership</td>
<td>PM1: I feel membership on MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PM2: I feel that I belong to certain groups on MB</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PM3: I feel that other members of MB groups treat me as a close friend</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PM4: I like other MB group members</td>
<td></td>
</tr>
</tbody>
</table>

*(continued)*
<table>
<thead>
<tr>
<th>Construct</th>
<th>Formative components</th>
<th>Measurement items</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived influence</td>
<td>PI1: I feel as if I am well known in the MB groups</td>
<td>Davis et al. (2002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PI2: I feel as if I have some control when MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PI3: I feel as if my MB postings are often reviewed by other members</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PI4: I feel as if the replies to my MB postings are frequent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microblogging addiction</td>
<td>DIC1: I often keep thinking about something I experienced on MB well after I have logged off</td>
<td>Davis et al. (2002)</td>
<td></td>
</tr>
<tr>
<td>Diminished</td>
<td>DIC2: when I am not using the platform, I often think about it</td>
<td></td>
<td></td>
</tr>
<tr>
<td>impulse control</td>
<td>DIC3: I cannot stop thinking about MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(DIC)</td>
<td>DIC4: I feel that I use MB more than I ought to</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DIC5: my dependency on MB usage sometimes seems beyond control</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DIC6: even though there are times when I would like to, I cannot cut down on my intentions to use MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loneliness/ depression (LD)</td>
<td>LD1: I am less lonely when I am using MB</td>
<td>Davis et al. (2002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LD2: I cannot see myself being without MB for too long</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LD3: MB is an important part of my life</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LD4: I feel helpless when I do not have access to MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social comfort (SCO)</td>
<td>SCO1: I am most comfortable when using MB</td>
<td>Davis et al. (2002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCO2: I receive more respect when MB than “in real life”</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCO3: MB relationships can be more fulfilling than offline ones</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCO4: I am at my best when I am MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCO5: I wish my friends and family knew how people regard me when I am MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCO6: when I am MB, I can be carefree</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCO7: I feel safest when I am using MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SCO8: I often find the use of MB peaceful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distraction (DIS)</td>
<td>DIS1: when I am on MB, I do not think about my responsibilities</td>
<td>Davis et al. (2002)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DIS2: through MB, I can forget the things that I must do but do not really want to do</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DIS3: through MB, I can ignore unpleasant things</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DIS4: through MB, I can procrastinate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DIS5: I feel that I use MB more when I have something else I am supposed to do</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table AI. Note: MB, Microblogging
### Appendix 2

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Items</th>
<th>Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social communication</strong></td>
<td>SE1</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>SE2</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>SE3</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>SI1</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>SI2</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>SI3</td>
<td>0.47</td>
</tr>
<tr>
<td><strong>Information</strong></td>
<td>IA1</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>IA2</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>IA3</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>IA4</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>IS1</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>IS2</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>IS3</td>
<td>0.77</td>
</tr>
<tr>
<td><strong>Entertainment</strong></td>
<td>LE1</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>LE2</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>LE3</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>LE4</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>CW1</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>CW2</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>CW3</td>
<td>0.85</td>
</tr>
<tr>
<td><strong>Content gratification</strong></td>
<td>C1</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>C3</td>
<td>0.92</td>
</tr>
<tr>
<td><strong>Process gratification</strong></td>
<td>P1</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>P2</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>0.94</td>
</tr>
<tr>
<td><strong>Social gratification</strong></td>
<td>SP1</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>SP2</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>SP3</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>SP4</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>PM1</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>PM2</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>PM3</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>PM4</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>PI1</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>PI2</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>PI3</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>PI4</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>Diminished impulse control</strong></td>
<td>DIC1</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>DIC2</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>DIC3</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>DIC4</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>DIC5</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>DIC6</td>
<td>0.76</td>
</tr>
<tr>
<td><strong>Loneliness/depression</strong></td>
<td>LD1</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>LD2</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>LD3</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>LD4</td>
<td>0.82</td>
</tr>
<tr>
<td><strong>Social comfort</strong></td>
<td>SCO1</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>SCO2</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Table AII. Results of the confirmatory factor analysis (continued)
### Table AII.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Items</th>
<th>Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SCO3</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>SCO4</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>SCO5</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>SCO6</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>SCO7</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>SCO8</td>
<td>0.72</td>
</tr>
<tr>
<td>Distraction</td>
<td>DIS1</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>DIS2</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>DIS3</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>DIS4</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>DIS5</td>
<td>0.55</td>
</tr>
</tbody>
</table>

**Corresponding author**

Wei Xu can be contacted at: weixu@ruc.edu.cn
Social media’s have-nots: an era of social disenfranchisement

Xinru Page
Department of Computer Information Systems, Bentley University, Waltham, Massachusetts, USA

Pamela Wisniewski
Department of Computer Science, University of Central Florida, Orlando, Florida, USA, and

Bart P. Knijnenburg and Moses Namara
Clemson University, Clemson, South Carolina, USA

Abstract

Purpose – The purpose of this paper is to investigate the motivations, concerns, benefits and consequences associated with non-use of social media. In doing so, it extends Wyatt’s commonly used taxonomy of non-use by identifying new dimensions in which to understand non-use of social media. This framework encompasses a previously unidentified category of non-use that is critical to understand in today’s social media environment.

Design/methodology/approach – This is an exploratory interview study with 17 self-identified social media non-users distributed across age groups and socioeconomic backgrounds. A thematic analysis is conducted based on a novel extension of Wyatt’s framework and the risk-benefits framework. This is supplemented by open coding to allow for emerging themes.

Findings – This paper provides empirical insights into a formerly uninvestigated population of non-users who are prevented from using social media because of social engagement (rather than functional) barriers. It identifies how these individuals face social consequences both on and off social media, resulting in social disenfranchisement.

Research limitations/implications – This is an initial exploration of the phenomenon using an interview study. For generalizability, future research should investigate non-use with a broader and random sample.

Practical implications – This paper includes design recommendations and implications for social media platform designers to mitigate the consequences experienced by socially disenfranchised non-users.

Social implications – Addressing concerns of this newly identified class of non-users is of utmost importance. As others are increasingly connected, these non-users are left behind and even ostracized – showing the dark sides of social media use and non-use.

Originality/value – This work identifies types of non-use of social media previously unrecognized in the literature.

Keywords Social media, Facebook, Non-use, Technology acceptance, Social engagement, Disenfranchisement

Paper type Research paper

Introduction

Social networking sites (SNSs) have experienced explosive growth over the last decade. Facebook, the most popular SNS in the USA, has been adopted by 79 percent of online Americans, while other platforms such as Instagram, Pinterest, LinkedIn and Twitter plateau at around 30 percent adoption (Greenwood et al., 2016). With about two-thirds of Facebook’s users logging in at least daily (Duggan and Smith, 2014), social media usage has become a norm. However, there are still many online adults who are not using social media. Most of the literature focuses on social media use and even overdose (Dong et al., 2018; Yang et al., 2016), but non-use is not nearly as well understood. This paper investigates the experiences of individuals who self-identify as “social media non-users,” examining the barriers and motivations for use and highlighting the tensions between choosing to engage in, vs abstain from, social media. Thus, this work deeply examines the dark sides of social media use and non-use from the perspective of those who have chosen to disengage.

Information systems (IS) researchers often examine non-use transitively through the lens of technology acceptance and adoption models such as the technology acceptance model (TAM).
and unified theory of acceptance and use of technology (UTAUT) (Davis et al., 1989; Venkatesh et al., 2003); as such, there is a wealth of research on factors that may lead to or detract from eventual adoption. While these models originate from the era before social media, they have been applied to social media adoption (Dhume et al., 2012; Lee et al., 2012; Rauniar et al., 2014; Wu and Chen, 2015). IS researchers have since moved beyond studying social media adoption to integrating post-adoption theories, such as the IS continuance model (Limayem et al., 2003), which helps explain sustained social media usage over time (Wu et al., 2018). Yet, we argue that social media non-use cannot be understood simply as the antithesis of social media adoption and use. While models like TAM, UTAUT and IS continuance may be able to capture the influence of functional and hedonic factors of adoption and use, they may miss the more nuanced social dynamics that are essential to holistically understanding the complete social media experience – or more specifically, the experience of both users and non-users (Baumer et al., 2013; Boyd and Ellison, 2007). In short, technology acceptance, adoption and continuation models tend to focus on the user, and we aim to investigate the experience of the non-user.

Our work is inspired by past research that highlighted the importance of studying non-use (Satchell and Dourish, 2009; Wyatt, 2003) and philosophically challenged the underlying assumption that technology acceptance and adoption should be the de facto standard in systems-based research. For example, Wyatt (2003) argued that when usage is considered a privilege or benefit, such as having internet access, non-use is erroneously equated to a lack of access or a state of deprivation (e.g. disenfranchisement). The idea that non-use could be voluntary is often ignored. Similarly, Satchell and Dourish (2009) pointed to an implied “utilitarian morality” where a “good” user is one that adopts the technology. Other work approaches non-use as a means to resist the temptation of overuse (e.g. addictive behaviors), regain a sense of control over technology or engage in active resistance as part of a socio-political statement (Baumer, Ames, Burrell, Brubaker and Dourish, 2015; Baumer, Guha, Quan, Mimno and Gay, 2015; Portwood-Stacer, 2013; Schoenebeck, 2014).

A commonly used taxonomy of non-users was developed by Wyatt (2003), which varies along two dimensions: temporality of adoption – whether the individual had previously been a user, and choice – whether intrinsic motivations or extrinsic constraints were the primary rationale for non-use. While many researchers have implicitly extended Wyatt’s taxonomy (Baumer et al., 2013; Baumer, Guha, Quan, Mimno and Gay, 2015; Portwood-Stacer, 2013; Schoenebeck, 2014), few researchers have focused on creating a deeper understanding of the social and cognitive processes of non-use.

Background literature

Non-use has been studied in the context of technology (Satchell and Dourish, 2009), the internet (Wyatt, 2003) and social media (Baumer et al., 2013; Lampe et al., 2013; Tufekci, 2008). We first introduce Wyatt’s seminal taxonomy of internet non-use, synthesize new perspectives from subsequent non-use literature and then provide a more in-depth review of social media non-use literature relevant to our own work.

Wyatt’s taxonomy of non-use

As mentioned in the introduction, Wyatt (2003) developed a taxonomy of non-users along the dimensions of temporality of adoption and choice. Individuals who never had the chance to adopt new technologies were considered excluded, while expelled non-users had their access revoked. Due to extrinsic barriers related to the digital divide, such as lack of infrastructure, digital literacy or socioeconomic status, excluded and expelled non-users have both been called “disenfranchised” users (Satchell and Dourish, 2009; Wyatt, 2003). Alternatively, resisters were characterized as those who chose not to adopt, while rejecters had previously adopted but chose to disengage of their own volition.
Empirical extensions of Wyatt’s framework

Other non-use researchers reference and have implicitly extended Wyatt’s taxonomy to include other types of empirically studied non-use but without explicitly returning to the theoretical underpinnings of Wyatt’s framework. For the most part, these new categories of non-use fit within Wyatt’s theoretical framework, such as resisters who are disenenchanted or disinterested in adopting new technologies (Satchell and Dourish, 2009; Tufekci, 2008), or rejecters who abstain, leave, quit or take a break from social media (Baumer et al., 2013; Baumer, Guha, Quan, Mimno and Gay, 2015; Portwood-Stacer, 2013; Schoenebeck, 2014). However, we identified four categories of non-use from the literature that were orthogonal to Wyatt’s taxonomy: relapers, limiters, displaced and laggards.

Baumer et al. (2013) identified relapers and limiters as unique types of non-users. Relapers are non-users who have stopped using a technology but ultimately return or “revert” back to being users, often due to addictive tendencies (Baumer, Guha, Quan, Mimno and Gay, 2015). In contrast, limiters may use various platforms within specific parameters or constraints. Displaced users are those who use technology indirectly as a service (Satchell and Dourish, 2009) or exhibit secondary use through others (Wyatt, 2003). The non-use category of laggards, which hail from diffusion theory, implies that non-users are simply future users who have not “yet” adopted (Satchell and Dourish, 2009). In terms of adoption, these non-users could potentially be users in the past, present and future. As such, the current consensus is that non-use should be represented along a continuum, rather than the binary adoption or non-adoption (Baumer et al., 2013; Brubaker et al., 2016; Wyatt, 2003). Further, non-use is not a “singular moment” but a temporal process, “involving layered social and technical acts” over time (Baumer, Guha, Quan, Mimno and Gay, 2015; Brubaker et al., 2016; Rainie et al., 2013). We incorporate these new categories of current and future (non-)use in our suggested extensions of Wyatt’s theoretical framework shown in Figure 1.

Social media non-use

For the most part, those who have studied social media non-use have focused on rejecters of a particular social media platform. For instance, Stieger et al. (2013) found that Facebook “quitters” had higher levels of privacy concern, were more addicted to the internet and were more conscientious than users. Lampe et al. (2013) found that older adults with higher levels
of social bonding capital are less likely to be Facebook users. Other researchers have also studied non-usage patterns and motivations on Facebook from different perspectives, e.g. “break-takers” (Baumer, Guha, Quan, Mimno and Gay, 2015; Rainie et al., 2013), “leavers,” “relapsers,” “limiters” (Baumer et al., 2013) and “abstainers” (Portwood-Stacer, 2013). Fewer studies have examined non-use on social media platforms beyond Facebook, such as taking a break from Twitter (Schoenebeck, 2014), leaving Grindr (Brubaker et al., 2016) or resisting adoption of location-sharing social networks (Page et al., 2013). A few scholars studied SNS non-use in general almost a decade ago (Hargittai, 2007; Tufekci, 2008) when there were significantly fewer popular sites, types of services and lower adoption rates. To our knowledge, none of the current work on social media non-use has revisited the phenomena of non-use more holistically across all platforms.

The social media non-use literature also focuses on the key motivations for why users have chosen to reject various social media. Motivations have ranged from boundary regulation or privacy concerns, data misuse, concerns about the amount of time spent on social media, communication style, banality or absence of compelling content, excessive gossip or drama, feeling judged, feeling manipulated, or active resistance as a political statement, to mere disinterest (Baumer et al., 2013; Baumer, Guha, Quan, Mimno and Gay, 2015; Lampe et al., 2013; Page et al., 2013; Portwood-Stacer, 2013; Rainie et al., 2013; Schoenebeck, 2014; Stieger et al., 2013). While the motivations for non-use across various platforms and contexts from prior studies have been insightful, they have moved away from Wyatt’s original framework and often overlapped. As a result, they lack a clear framework in which to ground their empirically driven findings. Our work suggests an extension to Wyatt’s framework to facilitate more cohesion as the body of non-use literature continues to grow.

Privacy calculus
Most of the existing research on social media non-use has undertaken initial exploration by identifying categories of non-use. Specifically, most work describes a taxonomy (Wyatt, 2003) or classifications (Baumer et al., 2013; Lampe et al., 2013) for different types of non-users. Our work moves beyond these descriptions by focusing on the social and cognitive processes that govern non-use transitions and practices. This leads us to map out users’ motivations, benefits, risks and concerns regarding the use or non-use of social media and integrate them into Wyatt’s framework. This novel contribution to non-use research is closely related to research on the privacy calculus (Laufer et al., 1973, Laufer and Wolfe, 1977), which refers to the cognitive process that underlies people’s decisions regarding the disclosure of personal information.

The privacy calculus can be seen as a privacy-specific instance of decision-making theories like utility maximization or expectancy-value theory (Awad and Krishnan, 2006; Li, 2012; Rust et al., 2002; Stone and Stone, 1990). These theories describe the process of evaluating and trading off positive and negative aspects of decision options (Bettman et al., 1998; Fishbein and Ajzen, 1975). When it comes to privacy decisions, these aspects are the perceived risk and perceived relevance of sharing a certain piece of information with a certain recipient (Featherman and Pavlou, 2003, p. 1,036; Li et al., 2011; Stone, 1981); whereas for use/non-use decisions, these aspects may include a broader set of motivations, benefits, risks and concerns. Furthermore, our work reveals that people’s non-use calculus goes beyond the traditional individualistic notions of controlling information flow. Rather, following the notion of networked privacy (Marwick and Boyd, 2014), the trade-off tensions of the non-use calculus are related to the social dynamics around how this information is used. Moreover, looking through the lens of privacy calculus led us to uncover the “dark side” of social media use/non-use by explicitly weighing the benefits vs the risks of social media engagement for non-users. As explained in our results, a key finding of this paper is that there are inherent drawbacks from making either decision, leading to sub-optimal outcomes whether individuals used social media or not.
Building upon the existing non-use literature

Focusing on the cognitive process of non-use, our research is motivated by a desire to map out users’ non-use process in terms of their motivations, benefits, risks, and concerns regarding the use or non-use of social media within (and beyond) Wyatt’s framework. This conceptualization is closely related to the idea of the privacy calculus (Laufer et al., 1973, Laufer and Wolfe, 1977). Our research focuses on social media non-use processes across all social media platforms. We interviewed 17 adults who self-identified as “social media non-users” and asked them to walk us through their previous use of social media (if any), their motivations for use and/or non-use and their past and current perceived and/or real benefits, risks, and concerns associated with social media use vs non-use. We also investigate whether and/or why they envisioned using social media in the future. Our research was driven by the following research questions:

RQ1. Are there aspects of social media that are unavoidable and push people to engage, even putting pressure on non-users?

RQ2. What are negative consequences and benefits of use and non-use, i.e. what dark sides of social media do participants mention as the main motivations behind the non-use process?

RQ3. What barriers of social media use do participants find hardest to resolve, i.e. what dark sides of social media remain unresolved, even for non-users?

As these research questions are buttressed in Wyatt’s (2003) taxonomy of non-use, we initially coded our interviews on Wyatt’s dimensions of the temporality of adoption (i.e. past use or no past use) and motivations for non-use (i.e. extrinsic vs intrinsic). Our analysis revealed a new dimension by which to characterize use and non-use motivations – social engagement vs functional use. This new dimension helped us frame several of our key findings in a more cohesive manner.

The first finding that motivated our social engagement/functional use distinction was our initial surprise that many of our self-proclaimed “non-users” actually used social media for work-related information-driven tasks and communications (i.e. functional purposes), while they did not adopt social media for purely social use (RQ1). In their minds, social media (non-)use was exclusively determined by whether they used it for social engagement purposes.

The second finding that motivated our social engagement/functional use distinction was the fact that all 17 of our participants abandoned using social media for social engagement purposes indefinitely. Social motivations that encouraged past use (e.g. interacting with friends) were overshadowed by negative social – not functional – consequences and concerns (e.g. false sense of community) associated with being on social media. Some said they would consider using social media in the future but for functional purposes only (RQ2).

Given these results, we used the social engagement/functional use distinction to articulate what we call the “non-use calculus” to demonstrate how there is a dark side to both social media use and non-use. The socially motivated dark side of non-use further gives rise to a new class of social media non-users – the socially disenfranchised: individuals who would like to reap the social benefits of social media but feel like they are being prevented from doing so due to emotional and social factors, which burden them even in their current state of non-use (RQ3). We discuss how social media platforms fall short of meeting the social needs of non-users and recommend ways to mitigate these effects so that non-use can be an empowered choice, as opposed to a state of social deprivation.

In studying non-use as a process, our goal was to extend the types of social media non-users studied beyond that of rejecters of a particular social media platform, and use (and suggest extensions to) Wyatt’s theoretical framework to unify the various motivations for social media non-use (both new ones as well as those studied thus far). To do this, we
recruited participants who identified as social media non-users across all social media platforms. This allowed us to cast a wider (more comprehensive) net than previous social media non-use studies in terms of our sampling method, empirical data collection, thematic analysis and theoretical development. In applying a theoretical lens to this novel population of social media non-users, we make the following unique contributions:

- Extended Wyatt’s non-use framework: we interpreted our data through the lens of Wyatt’s (2003) non-use framework combined with the categories that emerged in subsequent non-use research. This allowed us to make an important distinction between intrinsic motivations for non-use vs extrinsic constraints preventing and/or discouraging use. Acknowledging the temporality of non-use, we mapped participants’ motivations to episodic narratives – past use or non-use, current non-use, potential future use – to understand non-use as a complex temporal process. Combining the driving factors behind non-use studies and acceptance studies, we asked about motivations for current non-use as well as contextual information to understand possible motivations for past or future use. In doing this, we add to the cumulative understanding of non-use.

- Articulated the non-use calculus: drawing from the privacy calculus theory (Laufer et al., 1973, Laufer and Wolfe, 1977), we further unpacked participants’ motivations behind concerns regarding (potential or actual) negative consequences and (perceived or actual) benefits of use and non-use. This additional layer of analyses was not included in previous research on non-use and allowed us to discern the driving forces behind users’ use/non-use decision. This analysis also articulated the link between the non-use literature and the dark side literature by showing that social media non-use, as well as use, can often lead to unintended and socio-psychologically undesirable outcomes.

- Delineated between social engagement vs functional motivations and constraints: through our analysis, we found two types of usage motivation. The first was operationally driven functional use, which often had to do with disseminating and receiving information (flow of information). The second was desire for social engagement, which places emphasis on the interactions (influence of information). Similarly, the consequences and benefits of usage can be categorized as functional or social engagement issues or benefits. Another way to look at it is that functional goals and barriers are associated with one-way consumption or sharing, while social engagement goals and barriers have to do with two-way interactions and impacts. Within our sample, we found that social media non-use was mostly driven by social engagement issues since they were much harder for our participants to resolve than functional barriers. Ironically, these individuals continue to have social engagement problems arising from non-use. Acknowledging the important distinction between these two types of drivers, we further extend Wyatt’s (2003) taxonomy along this dimension to identify a new sub-class of expelled and excluded non-users, the socially disenfranchised.

**Methods**

**Data collection**

This study was part of a larger research project to understand people’s attitudes toward (and usage of, if any) social media. We conducted semi-structured interviews with individuals residing in the USA who were aged 18 or older. The criterion for being included in the current study was that participants self-identified as “social media non-users.” In our recruitment message, we gave examples of platforms considered social media by recent Pew Research Center studies (e.g. Facebook, Twitter, Instagram, Snapchat and LinkedIn) (Greenwood et al., 2016). Additionally, we drew on the social psychology literature that explains how
people's social and psychological needs change as they move through various developmental stages in life associated with age (Erikson, 1959; Newman and Newman, 2014). These needs influence their everyday behaviors. This has also been shown to affect social media perceptions and usage amongst social media users (Page and Marabelli, 2017). Given that our study is exploratory in nature, we wanted to capture a variety of perspectives and motivations. Thus, we took a stratified sampling approach to account for participants of different ages who might have different needs and behaviors. Participants were recruited from multiple sources: a mid-sized private university in the northeast, the local community and industry and the extended social networks of several researchers. We also relied on referrals from these sources to gain access to non-users, who proved harder-to-reach. Interviews focused on motivations for non-use, history of past usage, triggers that would lead to future usage, benefits and consequences of use or non-use, and associated attitudes, concerns and contextual factors. In summer 2016, we conducted 17 semi-structured interviews with non-users from a variety of socioeconomic backgrounds in the USA. Interviews were conducted in person, over the phone, or via Skype based on the participant’s preferred communication method. One participant was interviewed over e-mail, being uncomfortable with verbal exchanges. All non-users had access to computers and/or smart phones capable of supporting social media and expressed confidence in their technical and physical ability to use it. This study was approved by the university’s Internal Review Board.

Data analysis approach

We conducted a mixed methods qualitative analysis that was first theoretically grounded using Wyatt’s (2003) taxonomy of non-use, combined with any extensions we identified in the literature (Baumer et al., 2013; Satchell and Dourish, 2009). Specifically, we examined temporality of adoption by asking participants questions regarding past social media use, their present state of non-use and potential future use. Second, across these time periods (e.g. past, present and future), we inquired about their motivations behind their use/non-use decisions, classifying these as intrinsic needs or extrinsic constraints. This more positivist approach (Yin, 2013, 2015) ensured that our work was well-aligned with the theoretical underpinnings of past non-use literature. Yet, to develop new insights, we combined this inductive approach with a more interpretivist analysis that allowed for new themes and dimensions to deductively emerge from our empirical data. This hybrid qualitative approach strengthens our overall contributions by enabling us to identify both underlying mechanisms and causal patterns (Lin, 1998) related to the phenomenon of social media non-use.

Next, we further unpacked the motivations using the privacy calculus framework (Laufer et al., 1973, Laufer and Wolfe, 1977) that frames disclosure decisions as the result of analyzing risks vs benefits. This framework has also been used in prior work studying social media non-use (Baumer, Guha, Quan, Minmo and Gay, 2015; Lampe et al., 2013) and adoption (Xu et al., 2009), but we chose to extend it beyond exclusively privacy-related rationales toward a more inclusive “non-use calculus.” Motivations and non-use calculus factors (i.e. negative consequences and benefits, both potential and actual) were open coded based on participants’ responses across their states of use and non-use in the past, present and future. Two researchers independently coded the interviews and then worked together to resolve any coding conflicts.

Results

Participant characteristics

All interviewees identified as non-users prior to their interviews, but interestingly, we often discovered during the interview that a participant actually did use some social media platforms. However, their use on these platforms was functionally and operationally motivated (e.g. to share or receive work-related information, communicate within their organization, access services) rather than motivated by social engagement (e.g. to reconnect
with old friends, engage in social interactions, strengthen relationships with current friends). This helped us understand that identifying as a social media user is linked to using these platforms for social engagement motivations. After observing these different types of social engagement vs functional usage, we categorized our interviewees’ past, present and anticipated usage of social media by Wyatt’s (2003) dimension of intrinsic vs extrinsic motivations for (non-)use and also along our emerging dimension of social engagement vs functional use/non-use.

We also note that about half our participants were displaced non-users who gain access to social media occasionally and indirectly through family members who share information they feel would be of interest to the non-users. Half had tried using social media in the past for functional purposes, and half had tried it for social engagement purposes (this includes some overlap, since not all interviewees had used social media, and others used it for both social engagement and functional purposes). Moreover, the interviewees are spread across all four types of non-use in Wyatt’s (2003) taxonomy: excluded, expelled, rejecters and resisters. However, these non-use designations were due solely to social barriers of use rather than socioeconomic or other practical constraints. For instance, our interviewees all had access to home computers and even smart phones and data plans. They also did not express technical anxiety or barriers to use and were capable of installing and using social media (in fact, most had done so in the past). Instead, there were normative social pressures to be on, and subsequently, off social media. Our interviewees expressed that largely everyone around them (including strong ties such as spouses) used social media. Some of them still use social media occasionally but for exclusively functional purposes. Therefore, a participant labeled as an excluded non-user may still use social media on occasion for functional purposes, but the label reflects exclusion due to social engagement barriers such as social anxieties or bullying.

In this section, we describe our interviewees along these four non-use dimensions from our extended version of Wyatt’s taxonomy.

Resisters. Four of our non-users never used social media for intrinsically motivated reasons. They all perceived that social media generated too much useless information and they would rather keep things “simple.” Three of the resisters were also concerned about data privacy. Participant J, a cancer researcher in her forties, explained, “I have sort of this underlying suspicion of putting myself out there on the web like that […] I just really love having things simple and under my control.” So she felt strongly about the benefits of staying off social media. However, we discovered that she did experience drawbacks such as missing out on seeing pictures of friends’ families and being able to see their children grow up. Interviewee M, a college coach in her fifties, also expressed data privacy concerns but acknowledged having a LinkedIn account for functional purposes just so people can contact her. Individual K, a civil engineer in his 50s, further emphasized how social media information can be incomplete and misleading. He did not feel he had any business making inferences from people’s information, unless the person decides to tell him directly.

Interviewee B, a youth counselor in a juvenile prison in her early 20s, differed from the others in that she did not feel concerned about data privacy. She regularly interacted with her close friends and family and explained how she just never felt a need for the information on social media. However, her recent move across the country had left her out of touch: “I found out just a couple of days ago that one of my friends from back home had a miscarriage and lost her baby and she was really, really upset about it and posted something on Facebook. And I found out later and I felt really bad because I wasn’t able to console her, because I didn’t know about it.” Like many other intrinsically motivated non-users, she had the expectation that if something was important, she would find out somehow. This recent experience violated her long-held expectations; “She should have told me; I would have called.” This suggested to us that some non-users might be missing out on more than they tend to realize.
Rejecters. We interviewed four individuals who had used social media but had intrinsic motivations to reject it, i.e. they all felt it was a waste of time. Interviewee D (a man in his 30s, working in the legal field) had hoped that Facebook would help him “connect with some people that I haven’t seen for a long time.” However, he had misgivings about the authenticity of information posted and general data privacy concerns, and he “eventually decided to cancel the account, because I just felt like I wasn’t getting as much out of it […] I didn’t really talk to the other people that I knew very often and neither did they talk to me.” D did not get the social engagement that he was hoping for and felt social media was not worth his time.

Interviewee G (a business owner in his forties, working in the finance and food industries) also considered social media a waste of time but had several functionally driven uses for it. He had a Facebook account that was once required for him to apply for a job, and he kept a LinkedIn profile, stating that: “I really don’t use it as a social media site. I’m more there if people want to contact me. I really don’t look at the connections. I don’t contact anyone in LinkedIn.” He further explained that he might utilize Facebook for his business: “We’re toying with some ideas of social media from an advertising perspective,” but explained that “that’s a completely different use.” This illustrates how social media may in fact support both functional use and social engagement goals.

Participant H (a scientist in his 40s) went so far as to question the motives of people posting on social media. He joined MySpace and Facebook long ago when they were novel, but saw no value in things such as happy birthday posts that people performed only because they were prompted by Facebook or invitations to reconnect with those he had lost touch with. After many years, he deleted his personal accounts but still maintains a Facebook page for his lab to post work-related information.

Finally, interviewee O is an elementary school classroom aide in her 60s whose son thought she would enjoy keeping touch with people and so he set up her Facebook account. After the first day, she stopped using it, un-intrigued by all the “day to day stuff […] I have a headache. I don’t want to hear that. It turns me off completely. That’s a waste of my time.” When her husband started using the account and interacting with their friends, she made sure that he changed the account name to reflect both of their names so others would not misattribute his comments, which would sound uncharacteristic. She also reflected on how all around her “people are just looking at [social media] and you don’t talk as much because you know they’re interested in what’s happening on their phone.” As a result, even though O felt she had been able to keep away from social media, we observed that her relationships and interactions were still affected by others’ social media usage.

Excluded. We identified four individuals who were excluded from using social media but not for reasons commonly explored in the literature such as access to technology or the technical ability to use it. Rather, issues around social engagement were the main impediments. F, a government employee in her 30s, stayed away from social media to avoid the pressure to keep up with others and maintain a perfect profile: “I would want to make it the most informative, pretty page – and then I would spend more time on the computer than I already do and looking, ‘oh, what is everybody else doing,’ and it would suck me in.” Additionally, she hated the fact that others would be able to comment on her profile and feared that this would shape her online identity. In fact, when she entered law school, students had all automatically been signed up for a Facebook account. When she started getting friend request notifications in her e-mail, she said, “I was so livid I literally called Facebook headquarters and was like, I want you to take everything down!” Similarly, she had installed Instagram thinking it was a photo editor but canceled her account as soon as she realized it was social media. This illustrates that even though some social media support functional goals, they may be avoided by non-users who are unwilling to engage in social activities.
Participant P, who was in her 60s and working in multi-level sales, similarly called it “a death sentence” to keep up with people on social media. She explained that her and her husband’s success in sales made it so people constantly wanted to connect with them and talk to them about how they made it: “We have a network of about 270,000 people. We have been invited countless times to participate in Facebook and LinkedIn and different sites […] if we only got involved with 10% of that […] 27,000 people, there is no way that we could possibly keep up with that. It would be consuming, totally.” Both F and P felt they had to engage with others if they were on social media. Hence, they avoided the platforms, despite missing out on the social benefits. Participant I, a middle school teacher, expressed concern about how using social media would diminish communication and social skills: “It’s almost too much information, but it’s their choice. And I feel like it’s also taking away from actual face to face or even phone conversations where people don’t feel a need to talk anymore […] communication skills are just getting poorer and they don’t know how to have a conversation.” She further observed how students “don’t know how to resolve any sort of conflicts because none of it’s face to face.”

Interviewee N, a postal worker in his 60s, was excluded from social media for a different reason. He explains, “I grew up in a time where there were no computers, so all the bullying I received was physical.” Now, as an adult, he is civically active and once again, “I seem to be a target […] That is one reason I haven’t embraced the whole social media aspect.” Despite this, N revealed that he planned to join Facebook in the future, since he felt his community organization was really pushing it to help spread their message. Participants I and F also anticipated needing to join social media in the future for functional purposes. Once their children were old enough, they foresee signing up in order to monitor those accounts.

Expelled. Some of the most compelling descriptions came from the five participants who were expelled from social media. Like the excluded individuals, it was social barriers rather than economical or technical prowess that led to their abandoning social media. Participant A, a computer programmer in his 20s who chose to be interviewed through exchanging e-mail because of “my social anxiety” around talking on the phone, describes how “I’ve interacted with pretty much every form of social media and had most of them at one time or another” before deciding “social media is a plague […] A massive waste of time” that “lets people pretend to be connected to others when they aren’t really.” He explains how people “have ‘friends’ that they block (but don’t actually unfriend) because they actively dislike them but feel some sort of social obligation to have them as friends. I believe this creates a sense of false community.” The breaking point for him came with “the social anxiety” caused by “situations like, ‘This person I knew in high school sent me a friend request. I liked them then but haven’t talked to them in ten years. Do I accept them or not?’ [These are] ridiculously uncomfortable for me.” He would “leave them in pending requests forever, instead of just accepting or denying.”

As a consequence of quitting social media, interviewee A “gets a lot of flak for not being on Facebook.” His “friends created [an account] to troll me for refusing to be on Facebook myself. They’ll check ‘me’ into mildly embarrassing places and post things pretending to be me.” Furthermore, he is often excluded from social events: “I get, ‘What do you mean you didn’t know about the party? Oh yeah, I keep forgetting you aren’t on Facebook. Sorry about that’ a lot.” After canceling his MySpace, Facebook, Twitter, LinkedIn and Snapchat accounts, he only has “a throw-away account under a fake name” on Facebook to test coding integrations for work, and Tumbler and Reddit accounts for functional purposes.

Participant C, a dance major in her 20s, similarly felt the pains of social exclusion:

I have a friend group that I’ve been friends with for about a year now, and they’ve recently stopped inviting me to come to stuff, because they all have a Facebook page together and they all tell each other through Facebook. And so, they just either forget to invite me or don’t tell me outside of Facebook. So, I’ve sadly had to lose a few friends because of it too.
It was difficult for C to bring herself to use social media such as Facebook since “I was bullied a little bit in elementary school and middle school and so I decided on social media […] I didn’t want them to have an easy way to access me.” She had seen on social media that “people don’t have a filter anymore and they feel like they can put up whatever they want on social media. But they would never say it in real life to a person’s face. And I just think there’s a cowardice in that that I just find sickening.” So when her university’s ballroom dance company asked everyone to keep in touch through its Facebook page, “I tried to set one up for a day, and it just gave me this disgusted feeling that I just couldn’t do it. I couldn’t bring myself to go against everything I’ve said for years. I deleted it right then, because I had no desire to be a part of that.” When asked about simply not accepting friend requests from the people who had hurt her, C explained:

The friend request, initially, would probably spark some anxiety within me. Like, reopening the old wounds from middle school that I’ve healed and don’t want to necessarily remember. And even though they were unkind to me, I still have an issue being rude to somebody. I just want to put good out in the world. And I don’t want to hurt their feelings, because I wouldn’t want anybody to feel the way that that person made me feel. But at the same time, I just don’t want to be in contact with them, or anything remotely close to contact.

For C, there was no simple solution to the complex social dynamics surrounding use or non-use. Surprisingly, C had started using Instagram a few months ago, but for functional purposes to promote her dance shows to close family and friends: “I took a picture of our posters as soon as we get them, to really start promoting it.”

Participant L, who was in her 50s and had been on disability for several years, had a Facebook account that her youngest daughter created for her before moving off to college. This would allow her mom to keep in touch and reconnect with her college friends and her children, as well as those she had taught over the years through a youth program. However, reading about others proved too overwhelming for L: “They’re getting married or this or that. And it feels like I need to be doing something for them or helping them, and there’s just too many of them. Additionally, it was difficult seeing that others “are always doing something and I feel like I don’t have a life. Like everybody’s always on vacation […] and I don’t have a job. I can’t go anywhere. So, it kind of makes me feel a little more sad […] they’re all going on with their lives and I’m not part of their lives.” Nonetheless, the biggest challenge proved to be bullying she received from her in-laws. Her daughter had helped her set the account so that it was only viewable by friends, but L’s sisters-in-law still found out what she had posted:

My husband came home mad. “My sister’s mad. You put something on there.” And I’m like, “I didn’t put anything on there, except for, you know, who I am and what I like, or something […] It’s a control thing, you know? So they tell him what I can and can’t do and whatever, and he goes along with it. So, my daughter said, “You’re right. This is not going to be pretty for you, so let’s just get you off.”

Despite trying to limit who could see her posts, L was still unable to avoid the watching eyes of her in-laws.

Another expelled interviewee, E, an attorney in her thirties, had both Facebook and LinkedIn accounts. LinkedIn was only used for sharing information for work, but she had tried to use Facebook to connect socially. She described how several times she had “engaged [with Facebook] temporarily only to get overwhelmed by the amount of content and disengaged again.”

Participant Q, who was retired and in her 60s, finally signed up for Facebook because a friend asked her to connect. In hindsight she reflected, “I wasn’t doing it for the right reasons. I was doing it because I felt like, well, she’s my friend, maybe I should do this. Not because I wanted to be on Facebook and to go see all her daily activities and all her
However, she did not make it far beyond her first sign-in before deciding to quit. Q was wary about letting social media shape her personality. She explained how people rely on their social media networks for advice and let others tell them how to feel, rather than learning to be self-reliant:

[If] I have a bad day, I can say some four-letter words, throw something down, and I start working myself out of it [...] [If] I put it in writing, people are going to continue that bad moment by having it come back at me, reinforcing how bad it is [...] You don’t learn to rely on your own character [...] Where is the time where you can become yourself? [...] If you’re told how to do something all the time, and for years, you’re always going to look for instruction of how to do things.

Q felt that she would lose her personal character and independence if she joined social media. Ultimately, she could not bring herself to stay on it.

Motivations for use, non-use and potential future use
As illustrated in the previous section, many of our interviewees had engaged with social media in the past. Some had been motivated by the social engagement factors that represent the core functionality of most social media. For some, these social engagement goals were intrinsically motivated, such as wanting to interact with friends, or keeping in the loop about people’s lives. Other interviewees had been extrinsically motivated, nudged by close ties such as their children’s or close friends’ interest in their social well-being. However, the barriers that arose led all of them to ultimately drop this type of usage.

In contrast, other interviewees had been functionally motivated: i.e. they treated social media as a tool to obtain information and to accomplish more practical, often work-related goals. This functional use could similarly be intrinsically or extrinsically motivated. Several wanted to have accounts just so people could contact them, like a “yellow pages,” or to share a message for their organization. In other cases, functional use was imposed upon interviewees, e.g. as a requirement to use third-party services, or to communicate with members of vocational, religious or community organizations that played a vital role in their lives. However, extrinsically motivated functional use brought about a noticeable tension in those who had social barriers but felt obligated to use social media to accomplish functional tasks and goals. As a result, those intrinsically motivated were still using social media for functional reasons, while those extrinsically motivated had mostly stopped.

Some reasons our participants gave for disengaging from social media were similar to ones previously found in the literature when examining rejecters of social media. These were typically more functional concerns such as data privacy (from corporations, government, future employers, etc.) (Baumer, Guha, Quan, Mimno and Gay, 2015; Lampe et al., 2013; Stieger et al., 2013), being inundated with dramatic or inane posts (Baumer et al., 2013; Rainie et al., 2013) or concerns about social media being misleading (whether it is “fake news” or just that people’s opinions are not researched or fact-checked) (Baumer, Guha, Quan, Mimno and Gay, 2015). These reasons were expressed by non-users across all four categories of the Wyatt taxonomy. However, we identified several additional motivations that have not all been a focus in non-use literature. These motivations for non-use, which came predominantly from socially excluded and expelled non-users, were deeply rooted in how they negatively impacted non-users’ relationships and sense of identity. They arose from four different sources of anxiety that we describe here.

Being a good social media citizen. Several non-users felt that being on social media required them to keep up with everyone’s posts and to present a well-crafted presence. This level of commitment overwhelmed them. As E, F and L demonstrated, the maintenance work and active engagement are so taxing for non-users that they did not feel they could be on social media at all. They believed that doing a poor job at keeping up with Facebook could hurt relationships more than not engaging altogether. This is because being on social
media sets the expectation for them to read everyone’s posts and keep others up-to-date about themselves in return. For some, such as F, the drive to do a good job was also intrinsic. She felt external pressures and also recognized her own natural tendencies to do well.

Relying on a false sense of community. Several non-users hoped to connect with others on a deeper level but decided that social media relationships were too superficial. This came in different forms. D’s experience was that social media is not always conducive to meaningful interaction that would make him feel more connected. On the other end of the spectrum, A experienced how interacting with others and having social media “friends” was a hollow experience. He believed that the motivations behind those friendships might be inauthentic.

Being bullied. Several of our non-users were being bullied by others offline (C, L, N), and some hoped social media would be an outlet for support. However, they found that they could not escape bullying even on social media. Privacy controls for rejecting friend requests and controlling who sees what were not enough to help these individuals, as we saw for C and L. In fact, several of our non-user interviewees expressed how information can get to the wrong person despite their best technical efforts to prevent this from happening. Moreover, personal or social convention can prevent users from using privacy features, as was the case for C, who still wanted to spare the feelings of those who had bullied her. She found that being on social media would mean being connected with past bullies and her present-day critics and would just “be another way to hurt my self-esteem.”

Changing who I am. Non-users felt wary that social media would change who they are as a person. Others’ posts and reactions can shape one’s own personality and character (as Q worried). Another concern is that what others post about the non-user can come to represent his or her online identity (as F expressed). Likewise, some (e.g. A) felt that social media promotes gossiping, cyberstalking and other maladaptive traits. Many non-users also observed how people were less interested in engaging with those around them and expressed how they were “competing” with social media and people’s phones.

Strikingly, none of our participants expressed a desire to rejoin social media for social engagement-motivated reasons. Missing out on social connections was an anticipated trade-off of avoiding more negative aspects of social media engagement. Only intrinsic, functional motivations moved some of our interviewees to anticipate joining social media in the future (but with an expectation of negative social and emotional impact). Specifically, the anxiety around keeping a child safe overrode personal anxieties and reservations for F and I, who might use social media to monitor their children’s future accounts. And N’s plan to join social media was driven by his desire to help his organization share their strongly held beliefs, despite anticipating “negative responses” targeted personally at him. Functionally and intrinsically motivated use was the only persisting type of usage we observed in our interviewees, so it is unsurprising that this was also the only type of future usage anticipated by others. This observation reinforces how non-users are fairly perceptive about how potential future use could work for them.

The non-use calculus
In the introduction, we described our desire to map out users’ motivations, benefits, risks and concerns regarding the use or non-use of social media. To this effect, we introduce a framework that is analogous to the privacy calculus but with a focus on non-use aspects that go beyond the traditional perceived privacy risks and benefits. To signify this broadening of the privacy calculus framework, we dub ours a non-use calculus framework. We use it to understand how participants weighed the benefits and drawbacks associated with social media use and non-use, leading most to disengage from social media altogether and some to limit usage to functional purposes rather than social engagement.
Benefits of non-use outweighed those of use. Those who had previously used social media perceived no social engagement benefits and only a few functional benefits of use. Recognizing that social media is the new virtual phone book, several appreciated being contactable through their accounts. Some liked the ability to promote their organization or a cause, and others found social media valuable for being invited to events. However, the benefits of not worrying about having to continuously share information and providing persistent access to oneself were more numerous. All interviewees emphasized how much time they believed that they saved by being off social media, as the volume of information was simply too overwhelming for them to keep up. Another perceived benefit was not having to worry about the privacy of their personal data. Additionally, they felt that they did not have to worry about the authenticity of information or people – instead, they perused what they perceived as more reliable sources of opinions and news that they had found in the offline world.

The social engagement benefits of non-use were equally compelling. Many interviewees were relieved not to have the pressure of keeping their profile updated and keeping up with others’ posts (a pressure that was largely self-imposed). Not being criticized or judged online (or having to use it in fear of criticism or judgment) was a major benefit for non-users who had been bullied growing up, or who were still being bullied in the offline world. A common phrase that participants used to describe these benefits was that non-use made “life simple.” They claimed that they felt they could be their authentic selves and not be influenced by their social networks (at least not by the online variety).

Interestingly, it seems that most of the reported benefits of non-use were in fact avoided concerns with the use of social media. Indeed, non-users felt that all concerns and negative consequences they perceived while using social media were resolved by not using social media. Moreover, even those who never used social media voiced concerns about the prospect of using it that corresponded to actual consequences experienced by others who had. This leads us to believe that non-users who have never engaged with social media are still able to infer how using social media would affect them, since none of their concerns were unfounded.

Unresolved social consequences of non-use. Social engagement consequences of non-use, however, were more difficult to resolve. In fact, almost every interviewee who had a specific social engagement concern that drove them toward non-use had a corresponding social consequence related to being off social media. These social concerns and consequences of non-use were most often voiced by expelled and excluded non-users. They felt alienated from social media due to the social risks that outweighed the social engagement and functional benefits of use. However, they were now experiencing these negative consequences regardless – a lose-lose situation that has not been highlighted in the literature. These unresolved consequences of social media non-use include having one’s online identity shaped by others, social isolation through lost social connections and offline bullying.

Others are shaping my identity. People on social media tend to post about and tag non-users regardless of their absence. Consequently, non-users’ identities are being represented by others who are on social media. Even though non-users themselves avoid having their character shaped through social media, those around them are still engaging in maladaptive behavior such as gossiping and impulsive posting as a result of social media use. Other users are thus able to shape the non-user’s identity despite their abstention from social media. Worse, it is harder for non-users to detect and act upon such acts of disclosure and/or misrepresentation. Instances such as A’s fake Facebook profile illustrate how social pressures may leave people no choice about having an online presence and identity.

Losing a sense of community and social connections. Although being on social media may create a false sense of community, non-users often expressed losing social ties and closeness offline as well. Many have been left behind and lost friends who are building stronger bonds with one another via social media but are no longer in touch with the non-user. C’s story clearly
illustrates how building closeness through social media can come at the expense of pushing others away. Feeling left behind was particularly common amongst our expelled and excluded interviewees. The face-to-face time they hoped to gain by going off social media does not happen, as those around them are “sucked into their phones” and socializing on social media: “I have something else to compete with; not just the TV or the newspaper, but Facebook” (Interviewee O). When it comes to invitations for social events, they are often forgotten or told at the last minute. They also may not find out how others are doing, since social media has replaced other means of communicating social news (such as photo-sharing sites like Flickr).

Bullying offline. Non-users, who disengage from social media because they are worried about being bullied online, are often still bullied offline. Worse, the social support that could have benefited them in social media remains unrealized. Participant L expressed how she “probably would have stayed on there if it wasn’t for the harassment of my relatives […]”. However, now she does not have social media as a way to connect with others, but she still has “in-law issues on a daily basis, of them calling, harassing.”

In summary, we consistently observed that functional consequences could be resolved by leaving social media. On the other hand, when it came to social engagement consequences, non-users found themselves in a lose-lose situation where they had a similar problem regardless of whether they were on or off social media.

Surviving as a social media non-user in a world of users
We identified several conditions that allowed non-users to soften the drawbacks of being off social media. Many interviewees, especially expelled and excluded individuals, relied on secondary or displaced use through a close family member such as a spouse or child. Although the literature often discusses displaced use as extrinsically driven (Satchell and Dourish, 2009), we found that in our sample it was mostly an intrinsically motivated decision. These participants did not want to be on social media, and thus, decided to get the relevant highlights through an intermediary. This allowed them to keep up with major events while protecting themselves from the negative social consequences of being on social media. However, there was still an extrinsic constraint, in that their intermediary had to be willing to facilitate. This intermediary would usually be the one to initiate the interaction, e.g. by showing the non-user posts that they believed would be of interest: “Every once in a while, he will show me some of my old students or friends or something. He will say, do you want to see them? And he’ll show me once in a while.” Although interviewee L’s husband shared on occasion, this participant wished she knew much more about what her friends and former social connections are doing.

There were also interviewees who relied on their friends to contact them through alternative channels, such as e-mail or phone calls. Unfortunately, this approach was never 100 percent successful. Participants described being forgotten. Often they would get the invitation eventually over e-mail but too close to the event date to be able to make plans to attend. In terms of getting social updates, several interviewees had the attitude that people would call them if it was important. Yet relying on alternative forms of communication proved to be inadequate for most non-users, especially as certain alternative channels have become antiquated (e.g. birth announcements in the mail or e-mail). These individuals were left with no means of receiving social news. As a result, many expressed the feeling of being socially ostracized or isolated, which we identify as a new type of disenfranchised non-use: social disenfranchisement.

Discussion
The non-social use of social media
In this paper, we presented and analyzed several types of non-use. While many of these can be represented within Wyatt’s (2003) framework, our results compel us to expand the
framework based on previous literature and our own findings. First, we extend Wyatt’s framework to show that current and future adoption are necessary temporal dimensions to consider when studying non-use (Figure 1). In our study, for instance, many of our self-identified “non-users” actually used social media for functional purposes, while others met their functional needs indirectly through displaced social media use of others. We also include relaspers and limiters as types of current or future social media (non-)users in our new framework. This shows the complex interplay between use vs non-use and addictive vs restricted social media usage patterns as they relate to both intrinsic and extrinsic motivations of use. As more research addresses the dark side of social media in terms of the pathological behaviors and gratification of use (Andreassen et al., 2012; Blackwell et al., 2017; Chen et al., 2012; Song et al., 2004), as well as a wide range of less severe but nevertheless socio-psychologically undesirable phenomena (Mäntymäki and Islam, 2016), this classification will be useful for integrating non-use into these discussions. Finally, we include laggards in this new framework to illustrate that while this is a specific type of non-use, we should not assume that all social media non-users are potential future users who have just not adopted yet.

We also introduce the dimension of social engagement vs functional motivations and barriers to usage (Figure 2). Tufekci (2008) also differentiated between “social and non-social uses” of the internet. She observed that users of Facebook and Myspace were significantly more likely to use the expressive internet, “the practice and performance of technologically mediated sociality” (p. 547). This is in contrast to the instrumental internet, which involves “information seeking, knowledge gathering and commercial transactions,” (p. 548) and was utilized by both users and non-users. Our study also uncovers a dichotomy between usage types, but across somewhat different boundaries. For example, modern social media like Facebook are sometimes essential for instrumental purposes (Cao et al., 2016; Wu and Chen, 2015), thereby making it harder for users to disengage. Thus, we contribute to the literature by showing how usage patterns have evolved over the last decade as social media has become mainstream.

By partitioning social media non-use based on social and functional motivations and barriers, one can see that when it comes to the social side, existing literature has focused primarily on rejecters, i.e. users who chose to leave of their own volition (Baumer et al., 2013; Baumer, Guha, Quan, Mimmo and Gay, 2015; Brubaker et al., 2016; Lampe et al., 2013; Portwood-Stacer, 2013; Schoenebeck, 2014; Stieger et al., 2013). Meanwhile, the general non-use literature has discussed aspects of non-use in terms of the “digital divide” and disenfranchisement due to financial or infrastructural barriers (Satchell and Dourish, 2009; Wyatt, 2003). Quite noticeably, two octants have been previously overlooked. Past research.

Figure 2. Empirical and theoretical extensions to Wyatt’s non-use framework
tends to focus on excluded and expelled non-users based on intrinsic and extrinsic functional barriers, such as economic and technological limitations that contribute to “the digital divide” (Wyatt, 2003). Yet, individuals who do not engage with or disengage from social media due to social engagement barriers (i.e. social disenfranchisement) have not been explored in depth in the prior literature (Figure 2). Our study addresses this gap by looking at those who are barred from use by social barriers rather than technical or economic barriers. Our results show that such barriers include being bullied, worrying about one’s social identity, being overwhelmed by social connections and being disappointed by a false sense of community.

We map out these barriers using a non-use calculus framework, which builds upon the privacy calculus framework (Laufer et al., 1973, Laufer and Wolfe, 1977). While the privacy calculus has been used in prior work studying social media non-use (Baumer, Guha, Quan, Mimno and Gay, 2015; Lampe et al., 2013) and adoption (Xu et al., 2009), we chose to extend it beyond exclusively privacy-related rationales, hence the term non-use calculus.

The dark side of social media
Our non-use calculus analysis uncovers that extrinsic, social barriers are not completely avoided by disengaging from social media. Indeed, non-users are still being bullied, their online identities are still (and now exclusively) being shaped by others, and they still do not find a stronger sense of community – in fact, being off social media further erodes their social connections and perpetuates social isolation. Given this persistent social disenfranchisement, one may argue that when it comes to social media use, the social media have-nots will soon find out that they are in a lose-lose situation. This finding provides an interesting contrast to existing frameworks of technology adoption and acceptance (Davis et al., 1989; Venkatesh et al., 2003), which have traditionally assumed use and non-use to be the result of diametrically opposite motivations; a similar analysis can be made regarding the privacy calculus (Laufer et al., 1973; Laufer and Wolfe, 1977). Rather, the tension between use and non-use embodies a new paradox that (not unlike the privacy paradox (Norberg et al., 2007; Spiekermann et al., 2001) cannot be resolved through a rational trade-off between the benefits and drawbacks of engaging with others through social media. This apparent paradox reflects the findings of existing work on technology acceptance. For example, perceived ease of use – a core concept of TAMs – has been found to not have a significant influence on post-adoption discontinuance (Parthasarathy and Bhattacharjee, 1998). Moreover, the personal motivations (Turel, 2013) and system characteristics (Cenfetelli, 2004) that have been shown to influence discontinuance and non-use are distinctly different from those that have been established to influence adoption and use. Consequently, Cenfetelli (2004) argues that “inhibitors deserve independent investigation on the basis that they are fundamentally different in nature and effects from previously established positively oriented perceptions within the technology acceptance and user satisfaction literature” (p. 487).

Our work also suggests that the dark side of social media may be very different for non-users than for users. Not only do non-users have a different set of drawbacks and social barriers, our results show that the perceived benefits may be distinct for non-users when compared with those of social media users. For example, although our results highlight how saving time was a major benefit for non-users that feel it is inappropriate to ignore information coming in through social media, many social media users actually do cope with the volume of information that way (Wisniewski et al., 2012). And while several non-users were highly worried about having to provide personal data, many social media users feel no strong social pressure to post such information (Stutzman et al., 2013). Even the authenticity that non-users attributed to information and people in the offline world differs from those social media users who would rather describe traditional news sources as biased (Wang and Mark, 2013). These insights point to a more nuanced calculus when considering the dark side of social media non-use.
Limitations and future research

We conducted a qualitative analysis to detect nuanced trends and patterns in use and non-use across a relatively small sample of users in the USA. Had we studied a different demographic of non-users, such as those in economically or politically disenfranchised countries, we would have likely drawn much different conclusions. This warrants a larger, more generalizable study that might confirm these patterns as they relate to different demographics and motivations of social media (non-)users. As such, a web-based survey would allow us to collect a larger and more diverse sample of responses to build upon our current findings and make them more generalizable to a broader context.

We also suggest that those who want to confirm our results using either of these methods should consider the apparent fluidity of a person’s self-assessed status as a “non-user.” In our case, our participants did not self-identify as social media users, while there are others who likely have similar usage patterns (e.g. functional purposes only) that would. Therefore, inclusion and exclusion criteria for future work on social media use and non-use should be carefully considered. In our study, we found that self-proclaimed non-users were likely to still use social media for functional purposes. Thus, our inclusion of these participants was beneficial: had we limited participation to those who conformed to a stricter definition of non-use, then our uncovered functional use/social engagement distinction would have been less pronounced and our description of it less nuanced. Social non-users who are functional users may similarly provide important insights into the topic of non-use in future investigations. For example, a “netnography” of such users may demonstrate how these functional-only users avoid “getting sucked into” Facebook’s endless opportunities for social engagement.

In terms of future research directions, we found that the concerns of social media non-users often conflict with the very core of what constitutes “value” in a social network. However, non-use rarely mitigates those concerns. In fact, as social life increasingly expands online, the side effect is that non-users are increasingly left out. Therefore, rather than improving social networks as a means to engage non-users, our uncovered functional use/social engagement distinction would have been less pronounced and our description of it less nuanced. Social non-users who are functional users may similarly provide important insights into the topic of non-use in future investigations. For example, a “netnography” of such users may demonstrate how these functional-only users avoid “getting sucked into” Facebook’s endless opportunities for social engagement.

In terms of future research directions, we found that the concerns of social media non-users often conflict with the very core of what constitutes “value” in a social network. However, non-use rarely mitigates those concerns. In fact, as social life increasingly expands online, the side effect is that non-users are increasingly left out. Therefore, rather than improving social networks as a means to engage non-users, we argue that social network developers also have a responsibility to design and develop the non-user experience. Use and non-use are intertwined, and designers must start focusing beyond the user experience. In line with this argument, we provide several suggestions based on our analysis that we believe may mitigate some of the consequences of non-use:

- Provide a consume-only way to interact with social media: some non-users are mostly concerned with the production side of social media but would like to consume the content posted by others. Public, non-reciprocal forms of social media, such as Twitter, make this type of usage much easier. A “follow only” version of more closed networks like Facebook could allow similar consumption-only interactions.
- Provide mechanisms for redress: non-users are sometimes mentioned, or even misrepresented, by social media users. While users can monitor such events, and even set privacy settings for such mentions, it is particularly difficult for non-users to find out about and mitigate such events. In addition to a “consume-only” interaction paradigm, the social network could alert non-users (via another channel, such as e-mail) when they are mentioned in others’ conversations (similar to Google Search’s “Alerts” feature) and give them a means to communicate with the author in case of a transgression.
- Integrate non-use channels within social media platforms: social media interactions, such as Facebook events, are rapidly replacing more traditional, non-social media communication channels, such as e-mail invitations or even personal phone calls. Participants in our study complained about not receiving invitations, notifications and announcements or receiving them later than others. By allowing social media users to engage with non-users directly through their preferred channels
(e.g. e-mail, text messages), non-users could stay informed about upcoming events or important announcements without having to use social media. One excluded interviewee expressed disappointment when this type of feature disappeared from Facebook – it became much more difficult to be in the loop on social events.

At this point, these suggestions are based solely on the non-user experience, and future work needs to be done to validate if they would, indeed, mitigate some of the problems faced by non-users. Most of these suggestions would be quite feasible to implement, but it is up to designers and/or stakeholders of social media platforms whether they would be willing to tangibly address the needs of non-users. This further highlights the tension of how social media non-users may be marginalized as “deviants” if they do not eventually ascribe to be social media users.

Conclusion
As systems researchers, we may sometimes forget that non-use should be a viable option for individuals who choose not to partake in a certain technology. In the case of social media, however, we found that the pervasive socialization that occurs via social media has left some non-users socially disenfranchised and at a loss for solutions. This creates a new phenomenon that we consider the dark side of social media non-use. On the one hand, we could try to address these concerns through improved designs that mitigate some of the potential and real dark sides of social media. On the other hand, we could instead empower non-users in their decision to disengage from social media by helping them mitigate the social consequences of non-use. Our research highlights this paradigmatic shift – of not trying to “solve” non-use as an inherent problem – which is often overlooked in traditional systems-based research that has a vantage point of trying to increase acceptance and adoption. Lest we leave behind the social media non-user, non-use should be considered an integral part of design. Because designing for the user experience alters the experience of the non-user, we call on designers and developers to take on the responsibility of designing for the non-user experience.

References


**Corresponding author**

Xinru Page can be contacted at: xpage@bentley.edu

For instructions on how to order reprints of this article, please visit our website: [www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: permissions@emeraldinsight.com
Tolerating and managing extreme speech on social media

Brett G. Johnson
Department of Journalism Studies, Missouri School of Journalism,
University of Missouri, Columbia, Missouri, USA

Abstract

Purpose – This study seeks to understand the opinions of internet users toward extreme speech on social media platforms and their willingness to censor such speech. The purpose of this paper is to examine how norms of freedom of expression are changing in an online communication environment dominated by these platforms.

Design/methodology/approach – Four focus groups were conducted in this study. Participants needed to use at least one social media platform daily. Groups were homogeneous in terms of race and gender: African-American females, African-American males, white females and white males.

Findings – Participants in general did not report a strong willingness to censor extreme speech on social media platforms. Rather, they expressed apathy and cynicism toward both their own and social media companies’ ability to combat extreme speech and make online discourse more positive. Female participants tended to value the overall health of public discourse and protection of more vulnerable social media users on social media platforms. African-American female participants called for platforms to recognize a special duty to protect minority users, whom they saw as responsible for the platforms’ success.

Research limitations/implications – Focus groups are useful for providing exploratory rather than generalizable data. However, by increasing the understanding of how individuals define extreme speech on social media, these data can reveal how individuals rhetorically shape the social media platforms and interpret their role in democratic discourse.

Originality/value – This research takes the rich field of studying tolerance toward extreme speech to new territory: the online realm where public discourse (and especially extreme discourse) is hosted more and more.

Keywords Social networks, Abusive speech, Extreme speech, Internet censorship, Racism and sexism

Paper type Research paper

1. Introduction

Emboldened by the unscripted and often racist and sexist campaign of Donald Trump, members of the self-labeled “alt right” have taken to sites like Facebook and Twitter to express hateful messages and attack individuals who are non-white or female, or who express liberal political opinions. In response to this caustic rhetoric, politicians, pundits and the public have called upon the social media platforms that facilitate user-generated content (UGC) to do something to curb the proliferation of extreme speech. Such platforms include “sites and services that host public expression, store it on and serve it up from the cloud, organize access to it through search and recommendation, or install it onto mobile devices” (Gillespie, 2017, p. 254), a definition that includes sites such as Twitter, Facebook and YouTube. However, when platforms have taken remedial actions such as banning alt right sympathizers from using their services, they have received criticism for censoring speech that would otherwise be legally protected (particularly in the USA), despite the fact that these companies have no legal obligation to act as public forums and publish every bit of UGC on their platforms.

These events have highlighted a new reality for the place of extreme speech in public discourse. As public discourse has moved increasingly online (Nunziato, 2003; Gillespie, 2010), the platforms that facilitate UGC must develop and abide by policies for regulating extreme speech (Braun and Gillespie, 2011; DeNardis, 2014). These platforms have an incentive to police extreme speech strictly: doing so protects their ability to monetize networked communication and UGC through retaining a solid base of users whose attention they can sell to advertisers (Schaedel and Clement, 2010; Gillespie, 2017). As the power of these platforms to regulate speech grows (Peters and Johnson, 2016), both the meaning of freedom of expression...
in a world of networked communication and the notion of proper governance of extreme speech online are called into question. One of the best ways to understand this new reality is to study the opinions of individual users.

This study blends theories of freedom of expression with scholarship on tolerance toward extreme speech to create a framework for exploring how the concept of freedom of expression is changing in a networked communication environment. To explore this concept, the study uses focus groups to examine individuals’ opinions toward extreme UGC as well as toward the extralegal regulation of such UGC by platforms. The study also seeks to explore how participants’ race and gender factor into the formation of these opinions. Findings reveal the following trends: participants in general reported a sense of apathy and cynicism toward social media companies’ ability to combat extreme speech and make online discourse more positive. Female participants tended to view the overall health of public discourse and protection of more vulnerable social media users (e.g., children, women and racial minorities) as the ultimate goals of managing expression on social media platforms. African-American female participants called for social media platforms to recognize a special duty to protect minority users, whom they saw as responsible for the success of those platforms.

2. Theoretical background
This study takes the perspective that melding theories and methods from the disciplines of mass communication and law enhances “a thorough understanding of the concept and practice of freedom of expression” (Cohen and Gleason, 1990, p. 110). It seeks to address freedom of expression not merely as an abstract concept defined by normative theory, but rather as a lived experience that reflects and shapes the civic values of our democracy (Gillmor and Dennis, 1981). To that end, this literature review examines how traditional First Amendment theories justify legal tolerance toward extreme speech in US society, and it synthesizes those theories with work by social scientists to study the attitudes and behaviors of US citizens toward extreme speech. This synthesis is designed to direct the focus of the paper on the issue of how social media users tolerate and manage extreme speech online.

2.1 Extreme speech
The term extreme speech can “describe a wide variety of expression, such as [H]olocaust denial, extreme pornography, and speech inciting hatred or likely to provoke public disorder,” as well as speech that “is seen to be discriminatory or perpetuates discriminatory attitudes” (Rowbottom, 2009, p. 608). The definition of extreme speech depends on disparate and evolving social norms (Post, 2009), and thus pinning down a clear definition of the concept is no easy task. This study will use a working definition of extreme speech as speech that expresses hatred toward any individual or group based on a certain biological or cultural attribute or expression, or promotes, glorifies or incites acts of violence or terrorism. This definition excludes extreme pornographic content and extreme versions of commercial speech, such as ads selling drug paraphernalia. The study relies on a working definition since one of the study’s goals is to understand how individual internet users define the concepts of extreme speech on social networks. This definition is broad enough to encompass a wide variety of speech that could cause emotional harm to its targets (Smolla, 1992; Calvert, 1997), or could potentially lead to physical harm through incitement to violence (Lidsky, 2011).

Due to the First Amendment to the US Constitution, extreme speech receives a very high level of legal protection in the USA. Since the middle of the twentieth century, courts have interpreted the speech clause of the First Amendment (“Congress shall make no law […] abridging the freedom of speech”) as affording an exceptional negative right against attempts by the government to punish speech (Schauer, 2005). Courts have carved out only a few exceptions to these protections, including prohibitions against the most prurient forms
of pornography ("obscenity"), false advertisements, true threats and incitement to imminent lawless action (Smolla, 1992). Courts have held that the government may not punish racist, sexist, pornographic, violent or seditious speech, as such proscriptions would amount to unconstitutional viewpoint discrimination (Smolla, 1992). Many scholars and jurists have proposed theories for why extreme speech receives such exceptional protection in the USA. A common refrain is the “marketplace of ideas” theory, which holds that the ability of rational human beings to discern truth from falsity among the many messages (even extreme ones) put before them is more desirable than entrusting the government with making those decisions (Schauer, 2005). For John Stuart Mill (1859/2001), truth without competing falsity “is but one superstition the more, accidentally clinging to the words which enunciate a truth” (p. 34). Schauer (2005) argued this philosophy fits hand-in-glove with the general libertarian bent of American political philosophy, whereby government is distrusted in its capacity to choose which viewpoints – no matter how hateful or extreme – should or should not be accepted. The desired outcome is that speech on matters of public concern will emerge from the cacophony as the speech most beneficial to society. However, such speech can only be assured protection if extreme speech used by the fringes of society is also allowed.

Other scholars are less concerned with extreme speech’s position as a foil for truth as they are with the value of extreme speech to forge hardy democratic citizens. Bollinger (1986) argued that “society adds something important to its identity, [and] is significantly strengthened, by […] acts of extraordinary tolerance” (p. 9). Bollinger (1986) contended that individuals have a “deep and profound difficulty in controlling a desire to censor or suppress any difference of belief, opinion, or way of thinking” (p. 92), and exposure to extreme speech – through exceptional First Amendment protections for such speech – helps temper that innate desire. Post (1995) complemented Bollinger’s tolerance theory by arguing that extreme speech is protected in the USA to encourage citizens to critically engage with offensive, challenging and heretical ideas. He argued that a major goal of the First Amendment is to facilitate the “separation of public discourse from the domination of civility rules that define the identity of communities” (p. 177). For example, the community of people in America who promote racial harmony and multiculturalism must be prepared to engage with the crude and hateful rhetoric of the community of racists in America to be able to critically approach the problem of racism. Meanwhile, Emerson (1963) contended that the exceptional protections of the First Amendment act as a “safety valve” for society’s relationship with extremists: it allows extremist groups to let off steam in full view of the public rather than force them underground where their anger could build and, unmonitored by officials, these groups could plot to express that anger through violence rather than speech.

Individuals’ day-to-day experiences with extreme speech are changing. In today’s networked communication environment, extreme speech is brought closer to individuals than ever before, as their Facebook and Twitter feeds have become the home to various breeds of racism, sexism, homophobia, terrorist propaganda, personal abuse and depictions of graphic violence (Johnson, 2017a, b). In a networked communication environment, the mental and emotion harms associated with hate speech have the potential to morph into physical harms when they take the form of cyber-harassment or targeted abuse of an individual (Franks, 2011; Citron, 2014). The ability to speak anonymously on many social media platforms lowers the social costs of expressing extreme viewpoints online, thereby leading to an overall more caustic climate in online public discourse (Claessens et al., 2003; Malouf and Mullen, 2008). The changing nature and ubiquity of extreme speech on social networks make such speech more difficult to ignore or combat with counter-speech. Traditional theories used to justify the value of extreme speech in society, such as the marketplace of ideas, can become less palatable when the collective workings of human reason fail to lessen, let alone eliminate, the harms of extreme speech in
online public discourse. Meanwhile, as platforms that facilitate UGC become the de facto public forums of the networked communication era (Nunziato, 2003; Peters and Johnson, 2016), a lack of tolerance for extreme speech on these platforms could lead to a narrowing of acceptable public discourse on these platforms, leading that discourse to become stultified. Thus, it is important to explore whether individuals define the norms of freedom of expression differently in an online context than in an offline one, particularly as these norms pertain to extreme speech and its importance to public discourse.

2.2 Measuring tolerance toward extreme speech

Although the First Amendment grants legal protection to extreme speech, studies have revealed since the 1950s that many individuals disagree with the extent of that protection. This body of research is based on the following premise: “Although the notion of free speech is nearly irresistible to many people, certain specific examples of how other people may choose to enjoy that right may cause alarm and provoke intolerance among various segments of the public” (Andsager et al., 2004, p. 78). Yalof and Dautrich (2002) outlined two competing (yet not necessarily mutually exclusive) hypotheses to explain this distinction. Their inconsistency hypothesis holds that individuals, despite expressing strong support for civil liberties in general, will – out of ignorance or confusion – become more intolerant when the exercise of these liberties becomes controversial or appears harmful (Yalof and Dautrich, 2002, p. 58). Meanwhile, their qualification hypothesis holds that individuals who tend to support civil liberties in general will be selectively intolerant toward ideas that either offend or do not comport with their political beliefs (p. 59).

These hypotheses help form an understanding of what tolerance for freedom of expression means in the context of speech on social media platforms. Support in this context would not be a categorical statement framed in terms of rights (i.e. in reference to a state actor’s restraint from restricting speech), but rather a political statement geared toward discursively defining key attributes of these technologies (Gillespie, 2007, 2010). Inconsistent or qualified tolerance for speech on social media platforms therefore could reflect support among users for what Johnson (2017b) called an “aggregational” theory of freedom of expression. Under such a theory, the sheer ability to speak on social media platforms is privileged over the quality of the speech on those platforms or the transparent or consistent governance of that speech. The potential problem of users buying into such a theory is that they may become less likely to want to engage with controversial or extreme ideas on these platforms, leading to uncritical and stultified public discourse (Gillespie, 2017).

Research on tolerance toward extreme speech reveals that freedom of expression has always been an issue sensitive to the social and political context of the time. Stouffer (1955) found tepid public support for communist speech in the 1950s. Several studies (Gibson and Bingham, 1982; Sullivan et al., 1982; Immerwahr and Doble, 1982) found mixed support toward the speech rights of racists and anti-war protestors in the 1960s and 1970s. Andsager et al. (2004) found a high degree of polarization surrounding support for hate speech at the height of the so-called “political correctness” movement in the 1990s, which began with public universities’ attempts to implement codes proscribing hate speech and later morphed into a movement that sought to condemn any forms of language used by the powerful as a weapon against the oppressed. The social and political context of the present study involves the proliferation of extreme speech via social media platforms, which are regarded as very powerful media institutions with the ability (if not the desire) to shape public discourse (Peters and Johnson, 2016). This context is important because it addresses how individual users engage with extreme speech on social media at a time when such speech is permeating online public discourse (Meddaugh and Kay, 2009; Kwon and Gruzd, 2017). The present context is also important because it allows scholars to explore whether individuals believe these powerful platforms have a duty to more strictly manage extreme speech (Johnson, 2017a).
To properly explore these issues, scholars must understand social media platforms run the gamut from popular services like Facebook or Twitter that enjoy hundreds of millions of users to fringe services such as 4chan or Gab that have fewer (yet more active) users, and these platforms experience extreme speech differently. Facebook and Twitter face the dilemma of wanting to promote public discourse on their platforms while not having users leave the platforms due to lack of policing the most extreme or offensive discourse (Johnson, 2017b). Meanwhile, Gab and 4chan are home to users who will push the envelope to post the most extreme possible messages on their platforms (Peters and Johnson, 2016).

Although many of the early studies in this field have been atheoretical, Andsager et al. found that support for freedom of expression tended to be highest among the more affluent and educated males in society, leading the authors to theorize that “people who are most secure within society are most likely to support expressive rights” (p. 258). This theory comports with the normative conclusions of feminist legal scholars (MacKinnon, 1985) and critical race theorists (Lawrence, 1990) who have argued that First Amendment jurisprudence has tended to favor dominant groups in American society (namely, white males). Therefore, for the purposes of this study, it will be important to understand how individuals’ social identities as defined by race and gender (albeit in a limited fashion) inform their opinions toward extreme speech and how social media platforms manage it. For instance, a key issue will be whether women and minorities harbor different opinions than white males on these issues.

The studies cited above have shown that individuals tend to express intolerance toward certain specific examples of extreme speech. However, Lambe (2002) has pointed out that there is a conceptual distinction between attitudes toward certain types of speech and the willingness of people to censor that speech. Thus, Lambe sought to examine whether individuals might express their intolerance toward certain types of speech in terms of a willingness to censor it via legal action. Participants were asked for their level of support for the following responses to harmful speech: a law banning the speech; requiring that the speech be expressed only at certain times or in certain manners; no response; and active support for the speech (Lambe, 2002, p. 222). Lambe found that factors such as age, political ideology and gender tended to predict willingness to censor in certain circumstances, such as women being more willing to censor both hate speech and pornography (Lambe, 2002, 2004). Lambe’s work is important for its attempt at distinguishing attitude (public disdain toward speech) from behavior (taking steps to act on that disdain and censor the speech). This distinction is valuable when it comes to understanding how individuals might seek to regulate extreme speech online by using tools afforded them by social media platforms. In particular, individuals have the power to ask a social media platform to remove extreme speech by “flagging” it (Crawford and Gillespie, 2016), they can rally an online crowd to band together against an extreme message in an attempt to ridicule it (Johnson, 2016), or they can post a message countering extreme speech (Lambe, 2002). Users can also ignore extreme speech by selecting which speech they view on a regular basis (Stroud, 2008). Therefore, assessing tolerance toward extreme speech in a networked communication environment must take into account the presence of these tools and how users define the norms of using them within specific contexts.

3. Research questions
Based on the scholarship reviewed above, this study seeks to address the following research questions:

RQ1. How do individuals conceive of the norms of freedom of expression in networked communication?

RQ2a. What are individuals’ opinions toward extreme speech in networked communication?
4. Method

The present study relies on qualitative methods to answer these research questions. This approach is based on Lambe’s (2002) suggestion that understanding the free expression schemas of the average person might best be achieved through the use of qualitative interviews or focus groups (Lambe and Reineke, 2009). This study seeks to answer Lambe’s call by using focus groups.

The method of focus groups was chosen for two reasons. First, focus groups allow participants the ability to build off of each other’s responses, thereby offering the potential for rich conversation based on shared experiences (Lindlof and Taylor, 2011, p. 183). Second, focus groups give participants the opportunity to collectively construct a shared reality surrounding abstract concepts – such as attitudes, opinions or descriptions of behavior – in an efficient and observable manner for researchers (Lindlof and Taylor, 2011, p. 275). In the context of this study, these group dynamics can enhance discussion on complex concepts (e.g. freedom of expression in a networked communication environment) that may be experienced very differently by each individual participant, thereby opening up the possibility for synthesis and categorization of these concepts (Lin et al., 2015).

Four focus groups were conducted for this study. Participants across all four groups needed to use at least one social media platform daily, a choice that was made to ensure familiarity with the topics discussed in the focus groups. The four groups were homogeneous in terms of race and gender: African-American females, African-American males, white females and white males. It is acknowledged that extreme speech is not exclusively racist or sexist, African Americans are not the only racial minorities who are targets of racist speech, and many individuals experience gender in a way that is far more fluid than the binary set-up of this study. However, the choice to set up the study along these lines was made due to the fact that extreme speech on social media is often directed at African Americans and women, which is a continuation of a historical trend in the USA (Kang, 2000; Citron, 2014). The group of white females was put in place to identify possible distinctions between the experiences with extreme speech of women within the dominant racial group in the USA and those of women within the country’s historically most marginalized racial minority. The choice to make the groups homogeneous was made out of a desire for facilitating open and honest conversation about extreme speech and not alienating or creating conflict among any participants due to racial or gendered topics. Park et al. (2006) relied on a similar strategy in their study using focus groups to assess audience reactions to racial stereotypes depicted in the film Rush Hour 2. The research team relied on homogeneity to “create an atmosphere where participants would feel comfortable discussing potentially sensitive topics, such as racism” (Park et al., 2006, p. 165). Homogeneity also can facilitate deep investigation of differences along the lines of race and gender regarding attitudes toward freedom of expression that scholars have observed (Andsager et al., 2004; Lambe, 2002; Sullivan et al., 1982).

Participants were recruited using a database of potential subjects managed by a research laboratory at a large Midwestern public university. In a survey sent to potential subjects via e-mail, subjects self-reported their race, gender and whether they used social media daily. Participants self-reported their age upon arrival at the research facility. The focus groups were held at a university building in a roundtable setting, and participants were offered refreshments and compensated $25 for participating. An African-American female graduate
student moderated the African-American female and African-American male groups, and a white female graduate student moderated the white female group. The author, a white male, moderated the white male group. Matching the characteristics of the moderator as closely as possible to the characteristics of the participants in a focus group is seen as an effective strategy to establish a trusting relationship between moderator and participants (Park et al., 2006). Such trust is important for increasing the likelihood that participants will share their opinions freely (Whiteside and Hardin, 2011). A common script of 15 questions was used in each group, though moderators were instructed to follow the script in a semi-structured fashion, thereby allowing for fluid yet consistent discussion of key concepts. Participants were given broad questions to discuss rather than scenarios involving specific examples of extreme speech, thereby allowing participants to bring up definitions and scenarios organically. Sessions were video recorded.

The focus groups took place in late June 2016 at facilities operated by the same research facilitation laboratory that conducted the recruiting of participants. Each session lasted approximately 95 min. The African-American female group had seven participants, the African-American male group had ten, the white female group had ten and the white male group had eight. The difference in the size of groups is attributable to recruited participants not showing up for the sessions. The median age of each group skewed young: 24 years old for African-American females (range: 19-52); 23 years old for African-American males (range: 19–26); 30.5 years old for white females (range: 21–57); and 28 years old for white males (range: 19–57). Participants were well-educated, with all but one of the participants either currently enrolled as undergraduate students or having a BA degree or higher. One African-American female had an associate’s degree. Therefore, although education level and age may not be the predominant factors informing participants’ opinions and experiences regarding these research questions, the findings of this study cannot be completely separated from these attributes. The findings also cannot be completely separated from the relatively young ages of many of the participants.

The author transcribed each video, and then utilized a constant comparative method (Glaser and Strauss, 1967; Onwuegbuzie et al., 2009) to code responses as they appeared on the transcript. This method of coding occurs in three stages: open coding, whereby data are first broken up into smaller units; axial coding, in which these units are grouped into categories; and selective coding, whereby these categories are then grouped into major themes related to the research questions (Lindlof and Taylor, 2011, p. 252). In the results reported below, it is recognized that the comments of one or two individuals do not necessarily stand for the consensus view of an entire group (Greenbaum, 1998, p. 15). Therefore, the results are reported in a way that seeks to balance the goal of reporting general or prevailing opinions with the goal of not conflating predominant opinions with unanimous or even majority opinions. This balance is important because the dynamics of focus groups may lead participants to be less likely to deviate from prevailing opinions, potentially chilling dissent (Hollander, 2004; Onwuegbuzie et al., 2009). Thus, sentiments reported herein as the “general” opinion of one group (or across two or more groups) are those sentiments that multiple individuals expressed in a similar fashion, or that were met with expressions of agreement among fellow group members (e.g. head nodding, single-word assents, etc.). Proportions of participants who shared a certain opinion are reported when they are known. Individual comments are reported when those sentiments are representative of common themes, or when they represent opinions that challenge or dissent from common themes. Responses from individual participants are denoted herein using acronyms for each group (AAF for African-American females, AAM for African-American males, WF for white females and WM for white males) followed by a number that was assigned to each individual in each group when they were seated for the focus group.
5. Findings

5.1 Opinions on freedom of expression on social media platforms

On the topic of what “freedom of expression” meant to participants in the context of social media platforms, the prevailing trend across all four groups was that social media platforms were unique spaces for public discourse compared to traditional public forums and that only social media users’ “valid opinions” should be allowed on these platforms. However, the definition of a valid opinion varied among participants. One member of the white male group said he saw social media platforms as “American sounding boards” on which he could “take the pulse of the public,” and that these platforms should be tolerant of extreme speech so that people like him could get a sense of the range of opinions held by users (WM4). One African-American male participant argued that social media platforms may not be sites of serious public debate, but they should nonetheless be as open as possible for the sake of entertainment: “Whether it’s hardcore leftists or hardcore conservatives, I just want to see people duke it out. It’s entertaining!” (AAM5). One African-American female participant associated freedom of expression with self-fulfillment – echoing the theory put forth by Emerson (1963) that freedom of expression affords an individual the power to realize his or her full potential as a human being through “the development of ideas, […] mental exploration and […] the affirmation of self” (p. 879). This participant saw social media platforms less as an online forum for public discourse and more as the ideal place to unapologetically share one’s identity with the world:

Freedom of expression is just being who you are. If you live your whole life thinking, “Oh, I can’t say this because it’s going to offend such-and-such, it’s going to offend that person, it’s going to offend that person,” you’re going to be walking on eggshells your whole life and not truly be able to be free and express yourself the way you want to express yourself (AAF7).

Among the four groups, several themes emerged regarding how social media platforms alter the characteristics of speech from an offline context. First, attracting an audience is easier compared to speaking in a traditional public forum. As one white female participant put it, “Instead of going to a public space, where people just walk by and ignore you, you post it online because you have people who probably think similarly to you, […] so whether you’re right or not, you have some positive affirmation” (WF6). Second, individuals have the potential to reach a wide audience, one that is potentially global and consists of individuals the speaker did not intend to reach. One African-American female participant saw this potential as sufficient justification for more proactive regulation of extreme content, arguing that such regulation would stop such content before it goes viral and becomes more harmful (AAF3). Third, allowing users to speak with anonymity – a feature of sites such as Yik Yak and 4chan – gives users the ability to say extreme things with little risk of punishment. One white female participant expressed worry that such anonymity was leading online trolls to define the concept of freedom of speech as “free speech without consequences” (WF6). Finally, online content is permanent. Whereas an individual could once say something extreme in public with little risk of it being recorded and archived, what an individual says online could remain for a long time. Three participants in the African-American male group said that they would enhance this sense of permanence by taking screen shots of content that they found extreme, thereby preventing speakers from eliminating the speech from public view by deleting it. These indications are consistent with what many scholars have said about the potential of social media platforms to foster and exacerbate extreme speech (Kramarae and Kramer, 1995; Franks, 2011; Citron, 2014). However, one or two participants in each group noted that social media platforms can enhance the positive aspects of speech from an offline context, giving them the potential to improve public participation in democratic discourse. As one African-American female participant said, “That’s the point of why they have Twitter, so people can be engaged” (AAF5).
5.2 Opinions on extreme (and offensive) content

On the issue of how to define the concept of “extreme speech,” the prevailing trend across all four groups was that drawing distinct boundaries around extreme speech was a difficult task. Similarly, the prevailing viewpoint was that determining when speech becomes too extreme or too offensive to be allowable in online discourse was also a difficult task. Participants in each group tended to agree that “personal attacks” amounted to extreme content that always should be banned. However, within each group, participants could not agree on the definition of personal attacks. For some, a personal attack meant speech that actually attacked a specific individual. As one African-American male participant put it, “For me, for something to be offensive, it has to be like fighting words. If you’re not going to scrap over it, then it’s just [not offensive]” (AAM5). On the other hand, some participants (particularly within the African-American male and African-American female groups) regarded speech that targeted a group of people (e.g. all women, all homosexuals and all African Americans) as a possible personal attack. In other words, they believed that to target a group is to target an individual member of that group. “The things that I find offensive are anything that is like a personal attack on the community I’m from or the community that my friends are from,” one African-American male (AAM3) said. Four participants within the white male group indicated that incitement to violence was central to their definition of extreme speech, something the other groups did not mention. However, two of these white male participants disagreed with the other two on whether a message in support of a terrorist organization amounted to incitement or should be considered allowable political speech. The predominant view in the African-American female and the white female groups was that content that over-sexualized women and (especially) young girls should be considered extreme. One participant in the African-American male group also identified this type of content as extreme, though no other males did.

The prevailing trend among all four groups was to describe extreme and offensive speech in terms of the speaker’s intent, intransigence or use of inflammatory rhetoric more often than in terms of the content of the speaker’s message. However, two participants each from the African-American female and African-American male groups also expressed that sometimes offensiveness is unintentional, just a matter of ignorance (e.g. using coded racist language, such as the phrase “All Lives Matter,” without being aware of its offensiveness to many African Americans). Two or three participants within each group argued that individuals needed to be educated about the power their words have and that self-censoring extreme messages can be a good thing online. However, one white male participant held an outlying opinion: “I grew up in the ‘sticks and stones will break your bones’ era. […] So I should be able to say whatever the hell I want” (WM4). In fact, this participant argued that being able to express extreme speech on social media platforms could enhance discussion by giving it a sense of honesty and reality. For example, this participant said:

Maybe that first, knee-jerk reaction [inherent in online speech] is the most honest, whereas in real life you might tamp that down. […] If my gut reaction is to say, “You stupid bitch” there’s a part of that that’s an honest reaction, and knowing that’s out there has some value to me.

Most of what you’re talking about [with hate speech] I wouldn’t have a problem seeing on social media. I would take it with a grain of salt. So, someone doesn’t like LGBTQ people. So be it! That’s their opinion and they’re entitled to their opinion.

5.3 Opinions on how platforms manage extreme speech

Regarding RQ2b, two or three participants within each group expressed uneasiness over the fact that intermediaries had great power over users’ content. For example, two
African-American males said that these intermediaries could use this power to weaken the power of individuals to bring about social change through social media. One white male and one white female each saw the power of intermediaries as too arbitrarily enforced, allowing some types of extreme speech to persist despite the fact that users may have flagged them. Two members of the white male group said they worried that if intermediaries were too heavy-handed with their control of user content, some users would elect to take their extreme viewpoints to fringe, “anything goes” social media platforms such as 4chan or Gab, where they could become more extreme. Such a sentiment evokes Emerson’s (1963) “safety valve” rationale for tolerating extreme speech.

The prevailing trend across all four groups was an expression of cynicism toward social media platforms and their ability to regulate extreme speech. For example, one white female participant argued that platforms either were indifferent toward or tacitly accepted the extreme content posted by users. She saw this indifference as integral to these intermediaries’ business models:

The reason why these things [extreme speech] aren’t being monitored is because it stirs up, it creates traffic, it’s driving [...] their business [...] which generates dollars for them. That’s what it’s all about. Media want that (WF5).

This cynicism extended to opinions of platforms’ community standards. For one African-American male participant, “community standards are just there so things don’t get too crazy” (AAM5), rather than to set bright-line rules. One white male participant saw social media platforms trying to balance “the desire to have a free and open platform for communication with the desire to have a safe and welcoming platform for individuals of all inclinations” (WM2), and that community standards were a reflection of where these companies sought to place that balance. The prevailing belief within the white male and African-American male groups was that the market would regulate management of content; social media companies, they argued, would eventually understand which content should be removed and which should stay in order to prevent a major attrition of users. However, there was some disagreement within each of these groups about whether such market-driven regulation was a good thing. Two participants in the white male group expressed a fear for the possibility of increased fragmentation of social media platforms, with some catering to more extreme groups and others to mainstream society. Echoing the concerns of Gillespie (2017), these participants worried that such fragmentation would lead to a more banal and stultified public discourse taking place on mainstream platforms like Facebook and Twitter.

When asked for their opinions on the phenomenon of individuals banding together to pressure intermediaries to remove extreme content, most participants appeared to agree that this practice was predominantly positive, either in its ability to serve as a check on extreme discourse, or its ability to build solidarity around a particular cause. One African-American female participant noted that when she wanted to do something about an extreme post, she would post a message “to kind of spread the word, hoping that more people will spread it on and on, and that it will get to the right official” at the social media platform who could do something about it (AAF4). A fellow African-American female agreed: “If we can make a difference in numbers, then I think we should do that [band together] more” (AAF7). A white male participant framed banding together as a natural means of self-regulation, arguing that “people do a good job of regulating each other” online (WM5). One African-American male participant said that banding together gives him a sense of solidarity with others who are experiencing racism online by helping him recognize that “I’ve got brothers dealing with the same things I’m dealing with” (AAM3).

However, a couple of participants expressed negative views toward the phenomenon of banding together. These participants appeared to perceive banding together as a type of
online mob rule, and that this mob has little ability to distinguish between justifiable and unjustifiable actions against speech:

Groups banding together to get things removed counteracts the purpose of social media platforms. Platforms are there to have some sort of expression, you know, some sort of discussion. Removing that disrupts that whole system (AAM5).

There’s a mob mentality. If one person decides something is offensive, then they contact all of their friends and it just becomes this rolling wave (WF10).

5.4 Personally engaging with extreme content
Regarding RQ3, the prevailing trend across all four focus groups was for participants to express a reliance on multiple strategies for dealing with extreme speech on social networks, and that the context surrounding each encounter with such speech dictates which strategies they use. Strategies differed depending on the closeness of the relationship between the speaker and the participant. Two participants in the white female and African-American female groups each expressed willingness to “unfriend” (i.e. discard from their social media world) loosely connected social media contacts if they posted extreme content, or “unfollow” (i.e. ignore their presence on social media for a period of time) more closely connected contacts if they made such posts. However, two white females said they do not unfriend people because, as one of them put it, they “don’t want to have to deal with [the awkwardness]” of explaining why they unfriended the person (WF10). Two white female participants said that if the speaker of extreme content were a close friend or family member, they would talk to that person face-to-face about why their speech upset them. For example, one white female participant recalled talking to her father-in-law after he posted a comment on Facebook that she found offensive: “If my son ever hears you say that term out loud, you will never see him again. So if you ever want to hear the word ‘grandpa’ […] you’re going to keep that [opinion] to yourself” (WF10). This model of managing online discourse resembles a “Thanksgiving Dinner,” whereby users seek to manage the discourse of loved-ones in bounded settings for the sake of peaceful coexistence, rather than try to prevent them from speaking or shun them altogether.

The common response among all four groups was a sense of apathy over their ability to make online discourse less extreme. The prevailing belief was that people who post extreme content are unwilling to change their behavior/beliefs, or that responding to them will make them more extreme. For some, that apathy led them to ignore extreme content, and the vast majority of participants reported not feeling a personal sense of duty toward combating extreme speech. Indeed, this sense of apathy related to another common response among all four groups: the notion that self-preservation online is important. Whatever duty participants said they might have felt toward improving online public discourse, they seemed more likely to temper it with concerns for their own online reputation, their personal safety, their well-being and the health of their offline relationships with online friends. One African-American male described dealing with offensive content as a waste of time: “I just keep moving. […] All this childishness, I don’t need it” (AAM4). In general, participants across all four groups expressed a sense of apathy that users with extreme viewpoints could ever change. Thus, participants reported a preference for not engaging with extreme users rather than trying to stop these users’ speech.

5.5 Role of race and gender
The findings above show a relatively high degree of commonality among certain responses of all four groups, while a few differences in responses that happened to fall along racial or gender lines are also reported. However, several significant differences arose along the lines of participants’ race and gender that are worth reporting separately.
The prevailing opinion in both female groups was to see males (particular white, conservative and less educated males) as the primary perpetrators of extreme content. As one white female participant put it, she would be more likely to report extreme content to Facebook if the content was posted by “a man who clearly has not been exposed to anything” (WF4), referring to a less educated man. One participant each in both of the African-American groups defined freedom of expression along racial lines. As one African-American female participant put it, “[White people] love to say ‘freedom of expression.’ […] Freedom of expression could work if you’re white” (AAF2). Three of the white males appeared to agree with this sentiment, as expressed best by one participant:

It’s very easy for us all as white men to say, well, I can just take that hateful opinion with a grain of salt because most of it would not directly affect us. White and male […] we’ve already got two things that we’re not going to be hated on about on the internet (WM1).

Although most participants overall said that they routinely ignore extreme speech, one or two participants in each group said they would take action on such speech (e.g. comment on it, unfriend the speaker or flag the speech) if they felt compelled to do so. The African-American participants who felt this way framed such compulsion as stemming from attacks on their core identities as African Americans, which echoes Feinberg’s (1985) notion that the most offensive speech often involves attacks against things individuals find precious. The female participants who felt compelled to take action framed this compulsion in terms of an “ethic of care” (Steiner and Okrusch, 2006). They argued that extreme content should be proscribed because it threatened to harm vulnerable groups (e.g. children, women or LGBT users), and that it threatened to poison online discussion as a whole, thereby harming all audience members.

One participant each within the African-American female and male groups stated that the phenomenon of people banding to get UGC removed from a platform would be unacceptable if it prevented marginalized groups from getting their message heard on social media. However, as noted above, the prevailing viewpoint among African-American females and males was that users banding together to denounce extreme content should be considered a laudable act of social activism.

Two participants each within the African-American female and African-American male groups expressed mistrust in digital media companies’ general desire to look out for their communities, a sentiment that mirrored documented mistrust of traditional news media by many people of color due to routine stereotyped portrayals (Entman, 1994). Two African-American female participants went so far as to call on services like Twitter and Snapchat to realize that African Americans have been instrumental in the growth in the popularity of these platforms, and to honor this community by showing a greater duty toward policing racist content on their platforms. As one of the African-American females put it:

I understand it’s a business. The problem is, you’re not listening to all the people that make the business popular. […] Twitter and Snapchat, particularly in the black community, have become very popular, and they [these platforms] owe a sense of responsibility to us (AAF2).

6. Discussion

The findings of this study must be viewed within the context of how users “rhetorically shape” (Gillespie, 2007, p. 75) the nature of social media platforms and their role in facilitating free expression. Except for a few outliers, participants generally expressed intolerance for extreme speech on social media platforms. Participants defined extreme speech in terms of its potential to undermine the ability of platforms to facilitate good speech, as well as its ability to inconvenience users through its polluting presence in this networked space. This suggests that participants conceive of social media platforms (particularly mainstream ones like Facebook and Twitter) as spaces for public discourse
that are distinct from traditional offline public forums – not merely in terms of the physical differences between the two contexts, but in terms of the norms that govern public discourse within them. Meanwhile, the prevailing trend across all four groups was support for the idea that social networking platforms can empower individuals to bring about change or facilitate individual self-realization through free expression (Emerson, 1963, p. 879).

Responses from participants did not necessarily indicate an overwhelming willingness to censor extreme speech on social media platforms, either through individual or collective action. Rather, the prevalent preference was for participants to ignore extreme speech when they confronted it on social networking platforms out of a desire to preserve their own well-being. Those who reported having flagged extreme speech did so out of a concern for protecting the ability of others to use social networks as places of positive and meaningful expression, while those who reported a tendency to ignore extreme speech did so out of a concern for their own well-being. These sentiments appear to run parallel to the two hypotheses put forth by Yalof and Dautrich (2002), whereby individuals tend to support freedom of expression in general but are inconsistently or strategically intolerant toward specific examples of extreme speech.

The prevailing trend among female and African-American participants was an expression of stronger opposition to the presence of extreme speech on social media platforms – much stronger than that from the participants in the white male group. These findings appear to align with the theory put forth by Andsager et al. (2004) that those in positions of social power and privilege (namely, white males) are more likely to be tolerant toward extreme speech. Female participants’ intolerance toward extreme speech on social media platforms appeared to come from a desire to protect the health of online discourse as a whole and protect the well-being of the most vulnerable users of social media platforms. These responses expand upon Lambe’s (2002, 2004) findings that women tend to show a greater intolerance toward extreme speech; it is not necessarily the case that women are more intolerant toward extreme speech because they are more in favor of censorship, but rather because they interpret the harms of such speech on a broader and more social level.

Some of the disagreements within and between the focus groups from this study are indicative of a politicization of the notion of freedom of expression, which continues a trend identified in previous studies on tolerance toward extreme speech (Lambe, 2002; Gibson and Bingham, 1982; Sullivan et al., 1982). One of the white male participants (WM4), who, as reported above, saw value in allowing many types of extreme speech on social media platforms, self-identified as politically conservative during discussion. Although WM4 was an outlier among the more liberal members of all four of the groups, his opinions are by no means outliers in society. Politicization of free speech is nothing new: some conservatives have long denounced the burning of the American flag, which the US Supreme Court has deemed protected speech; some liberals, particularly over the last few decades, have demanded that speech that offends vulnerable populations be banned in settings such as public university campuses (Andsager et al., 2004). However, politicization of freedom of expression at a time when extreme speech proliferates on social media could calcify the differences between those who defend extreme speech and those who seek to rid social media platforms of such speech, leading neither group to critically engage with the speech and come to an understanding of the very real social and political evils behind it. This outcome is concerning, as it could lead to a weakening of public tolerance for the exceptional protections of the First Amendment (Gajda, 2015). Future research should continue to explore the potential effects that politicization of freedom of expression may be having on the health of public discourse.

The prevailing trend across all four groups was an expressed desire for social media companies to do more to manage extreme speech on their platforms. However, a lack of trust in the ability of social media platforms to manage such speech was an equally
prevailing opinion. This mistrust was present among participants who believed these companies should take a more proactive approach to policing such content, as well as those who believed that these companies should do more to protect good speech, such as content posted by political activists. A growing body of scholarship has taken a critical stance against the lack of transparency of social media platforms in how they manage various types of content (Tufekci, 2014; DeNardis, 2014; Peters and Johnson, 2016). These companies continue to build their business on promising individuals the potential to publish all sorts of expression before wide audiences (Schaedel and Clement, 2010). The sheer ability of users to speak on these platforms is privileged over the promotion of a robust public discourse or the transparent governance of extreme or controversial speech (Johnson, 2017b). Based on the responses of the participants in this study, social media platforms would be wise to accept a duty to be transparent about how they manage UGC, and to follow clear standards aimed at both promoting free expression and preventing harm against users. Furthermore, to honor a special duty toward protecting non-white users that one African-American female participant called for, these platforms should consider the variety of social experiences of their users (especially women and minorities) when crafting their politics for managing extreme speech.

7. Conclusion, limitations and implications

The opinions of these focus group participants do not necessarily represent a snapshot of American public opinion on the issues of extreme speech and the ways in which social media platforms regulate that speech. The findings of this study must be tempered by the fact that participants across all four groups were relatively young, well-educated and politically liberal. Also, although homogeneity is a recommended feature of focus groups (Brennen, 2013, p. 60), homogeneity does not guarantee that participants’ answers are more valid or truthful than if the groups were heterogeneous, as “the social contexts of focus groups – [such as] the larger social structures within which the [group] discussion takes place – affect the data that are generated” (Hollander 2004, p. 604). Although it is possible that group homogeneity led participants in this study to respond to questions less truthfully, this risk was deemed smaller than the risk of alienating participants in a heterogeneous setting. To compensate for these limitations, outlying opinions that could reasonably be perceived as having greater weight in society at large were reported here as such.

Despite these limitations, the findings of this study offer fertile ground for future study to help scholars understand how social media users define freedom of expression, and how these definitions can be used to critically evaluate traditional First Amendment theories. These findings also offer helpful insight for social media companies eager to better understand and manage the problems posed by extreme speech on their platforms. For instance, these companies would do well to put a primacy on expression that has the ability to bring about positive change in society rather than base their business models on the “aggregational” value of freedom of expression (i.e. that the potential for users to speak on their platforms is all that matters; see Johnson, 2017b). Such a policy could allow extreme speech to remain on platforms when it has such instrumental value, but could trigger the justifiable removal of extreme speech when it does nothing but attack individuals or protected groups. Social networking sites should communicate this policy clearly to their users. They should also frame their speech policies by acknowledging that they have a duty to promote a healthy public discourse (particularly for minority users).

Acknowledgment

Funding: grants from the University of Minnesota Mixed-Method Interdisciplinary Graduate Group and the Association for Education in Journalism and Mass Communication 2016
Emerging Scholars Program provided funding for this research. The author would like to thank Neeley Current, Kenneth Haggerty and Shahzaade Cannon for their valuable assistance in facilitating this research. The author would also like to thank Rachel Grant and Marina Hendricks for moderating three of the focus groups in this research.

References


DeNardis, L. (2014), The Global War for Internet Governance, Yale University Press, New Haven, CT.


**About the author**

Brett G. Johnson (PhD, University of Minnesota) is Assistant Professor of Journalism Studies in the Missouri School of Journalism where he teaches courses in journalism and mass communication law. Johnson employs traditional legal research methods as well as mixed- and multi-method approaches to study the social impacts of the rapidly changing legal and technological regimes governing emerging forms of journalism and online communication. Johnson’s research is grounded in a strong appreciation for First Amendment theory, which allows him to assess the fluid and oft-conflicting interplay between the law and social norms. Brett G. Johnson can be contacted at: johnsonbg@missouri.edu

For instructions on how to order reprints of this article, please visit our website: [www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: permissions@emeraldinsight.com
The dark side of news community forums: opinion manipulation trolls

Todor Mihaylov  
Institute for Computational Linguistics,  
Heidelberg University, Heidelberg, Germany

Tsvetomila Mihaylova  
Instituto de Telecomunicações, Lisboa, Portugal

Preslav Nakov  
Department of Arabic Language Technologies,  
Qatar Computing Research Institute, Hamad bin Khalifa University, Doha, Qatar

Lluís Màrquez  
Qatar Computing Research Institute, Hamad bin Khalifa University, Doha, Qatar

Georgi D. Georgiev  
Department of Innovation and Consulting, Ontotext AD, Sofia, Bulgaria, and

Ivan Kolev Koychev  
Faculty of Mathematics and Informatics,  
University of Sofia “St. Kliment Ohridski”, Sofia, Bulgaria

Abstract

Purpose – The purpose of this paper is to explore the dark side of news community forums: the proliferation of opinion manipulation trolls. In particular, it explores the idea that a user who is called a troll by several people is likely to be one. It further demonstrates the utility of this idea for detecting accused and paid opinion manipulation trolls and their comments as well as for predicting the credibility of comments in news community forums.

Design/methodology/approach – The authors are aiming to build a classifier to distinguish trolls vs regular users. Unfortunately, it is not easy to get reliable training data. The authors solve this issue pragmatically: the authors assume that a user who is called a troll by several people is likely to be such, which are called accused trolls. Based on this assumption and on leaked reports about actual paid opinion manipulation trolls, the authors build a classifier to distinguish trolls vs regular users.

Findings – The authors compare the profiles of paid trolls vs accused trolls vs non-trolls, and show that a classifier trained to distinguish accused trolls from non-trolls does quite well also at telling apart paid trolls from non-trolls. But it achieves the best performance for users with a sizable number of comments in the forum, e.g. 100 or more. Yet, there is not such a limitation for troll comment detection.

Practical implications – The approach would help forum moderators in their work, by pointing them to the most suspicious users and comments. It would be also useful to investigative journalists who want to find paid opinion manipulation trolls.

Social implications – The authors can offer a better experience to online users by filtering out opinion manipulation trolls and their comments.

Originality/value – The authors propose a novel approach for finding paid opinion manipulation trolls and their posts.

Keywords Credibility, Community networks, Astroturfing, Sockpuppets, Trollness, Trolls

Paper type Research paper

1. Introduction

The emergence of user forums in electronic news media has given readers the opportunity to express their opinion, which has made such forums much more attractive to online users: many of them enjoy the discussion more than the actual articles. This is the bright side of web...
forums (Napoles et al., 2017). However, there is also a dark side (Moqbel and Kock, 2018): the perceived anonymity has given rise to the proliferation of users posting malicious comments (Cheng et al., 2017), often aiming to manipulate other users' opinion (Ma et al., 2017).

The practice of opinion manipulation has been a reality since the rise of community forums. It has been shown that user opinions about products, companies and politicians can be influenced by other users' opinions (Dellarocas, 2006). Thus, companies, governments and political parties started hiring people to write in social media from fake profiles (Ratkiewicz et al., 2011; Chen et al., 2013; Kumar et al., 2017). We call such people opinion manipulation trolls.

Our primary goal here is to develop automatic methods for exposing paid opinion manipulation trolls and their comments in news community forums. This would help moderators in their work, and potentially also investigative journalists, but above all, the regular forum users, who would be able to read forums without paid propaganda.

Finding opinion manipulation trolls automatically is a hard task, as there is no easy way to recognize or even to define precisely what they are; this also makes it hard to get training and testing data for training a supervised machine learning system. We approach the problem pragmatically: we assume that a user who is called a troll by several people is likely to be one. This simple operational definition allows us to collect training data about accused trolls and their comments, and to train classifiers to automatically distinguish them from the non-troll users.

This work is organized as follows: Section 2 describes the related work. Section 3 presents the hypotheses we investigate. Section 4 gives an overview of our approach. Section 5 targets accused troll users, Section 6 focuses on paid troll users and Section 7 studies troll comments. Section 8 studies the credibility of posts by accused troll users. Finally, Section 9 discusses theoretical and practical implications, limitations and future work.

2. Related work

Social media forums are a useful way for online users to share opinions (Chen, 2017), to obtain information about products and services (Lin and Xu, 2017; Bijmolt et al., 2014), as well as to discuss political issues (Kwon and Gruzd, 2017). They are also an essential channel for word-of-mouth marketing (Jalilvand and Samiei, 2012; Gvili and Levy, 2016; Chen et al., 2017; Choi et al., 2017; Ortiz et al., 2017; Xun and Guo, 2017).

However, the promise of social media to democratize content creation has also been accompanied by many malicious attempts to spread misleading information over this new medium, which quickly got populated by sockpuppets (Kumar et al., 2017), internet water army (Chen et al., 2013), astroturfers (Ratkiewicz et al., 2011) and seminar users (Darwish et al., 2017). Several studies have shown that trust is an important factor in online relationships (Ho et al., 2012; Ku, 2012; Hsu et al., 2014; Elbeltagi and Agag, 2016; Ha et al., 2016), but building trust is a long-time process and our understanding of it is still in its infancy (Salo and Karjaluoto, 2007). This makes it easy for politicians and companies to manipulate user opinions in community forums (Dellarocas, 2006; Li et al., 2016; Zhuang et al., 2018).

Community forums saw the proliferation of fake news and clickbait (Karadzhov, Gencheva, Nakov and Koychev, 2017; Mohtarami et al., 2018), aggressiveness (Moore et al., 2012) and trolling (Cole, 2015). The latter often is understood to concern malicious online behavior that is intended to disrupt interactions, to aggravate interacting partners and to lure them into fruitless argumentation in order to disrupt online interactions and communication (Chen et al., 2013). Here we are interested in detecting not just any trolls, but such that engage in opinion manipulation. This latter definition has also become prominent in the general public discourse recently.

Troll detection has been addressed using semantic analysis (Cambria et al., 2010), domain-adapted sentiment analysis (Seab et al., 2015), various lexico-syntactic features about user writing style and structure (Chen et al., 2012), as well as graph-based approaches over signed social networks (Kumar et al., 2014). There have been also studies on general...
troll behavior (Herring et al., 2002; Buckels et al., 2014), cyber-bullying (Galán-García et al., 2014; Sarna and Bhatia, 2017; Wong et al., 2018; Sezer et al., 2015), as well as on linking fake troll profiles to real users (Galán-García et al., 2014). Unlike that work, here we compare paid vs accused trolls vs non-trolls, and we develop automatic techniques to distinguish them. We further apply trolling as a feature for predicting the credibility of comments in news community forums.

Trustworthiness and veracity analytics of online statements is an emerging research direction, especially given the recent interest in fake news (Lazer et al., 2018). This is related to trollness, as trolls often engage in opinion manipulation and rumor spreading (Vosoughi et al., 2018). Research topics include predicting the credibility of information in social media (Ma et al., 2016; Mitra et al., 2017; Karadzhov, Nakov, Márquez, Barrón-Cedeño and Koychev, 2017; Popat et al., 2017) and political debates (Hassan et al., 2015; Gencheva et al., 2017; Jaradat et al., 2018), and credibility and factuality in community forums (Mukherjee and Weikum, 2015; Hardalov et al., 2016; Mihaylova et al., 2018). However, none of this work has used accusations between forum users of being a troll as a feature to predict credibility.

3. Development of hypotheses

Note that it is not a priori clear whether trolls should be easily identifiable as they typically abide by the formal forum rules and they try to leave an impression of being good members of the community, which earns them respect and consequently increases the impact of their trolling actions (Donath, 1999).

Yet, many forum users have a good sense of who behaves like a forum troll. In particular, it has been previously observed that labeling a forum user as a troll is a common way of accusing them of improper behavior (Gazan, 2016). However, there were no studies whether such accused users really behave like trolls (e.g. known paid trolls) or even whether they behave differently from regular users. This is the research gap that we want to bridge. We are further interested in the nature of the potential differences in behavior. Thus, we formulate the following hypothesis:

H1. A user who is called a troll by several people has a markedly different behavior from regular users across a number of characteristics.

Next, we want to study accused trolls vs opinion manipulation trolls in order to see whether we can automatically identify the opinion manipulation trolls using data about accused trolls. We further want to compare the profiles of accused trolls to profiles of known paid opinion manipulation trolls. Such a comparison has not been performed before as we are the first to study accused trolls. Previous work (Chen et al., 2013) has focused on suspected paid trolls rather than on confirmed ones (as we do here).

Overall, we compare the profiles of paid trolls vs accused trolls vs non-trolls, and we study whether data about accused trolls generalize well to paid opinion manipulation trolls, thus serving as a kind of distant supervision or noisy training data for the class we are really interested in (i.e. paid trolls). This gives rise to the following hypothesis:

H2. Paid opinion manipulation trolls behave like accused trolls, and it is possible to distinguish paid trolls vs regular users automatically.

Trollness has been previously addressed mostly at the user level. One notable exception is the work of Cambria et al. (2010), which targets troll comments. This is critical for early detection of the comments by new troll users as trolls often register new accounts in the forum, e.g. because of having been discovered by the forum users or by the forum moderator. In the case of newly-registered users, detecting a troll user based on accusations is problematic as it takes time before a user accumulates enough posts in order to have a chance to be called a troll by several other users in the forum or for software to have enough
user activity in order to be able to do reliable automatic classification. Thus, we also study whether we can distinguish comments by accused/paid trolls vs such by non-trolls, which yields another hypothesis:

\[ \text{H3. It is possible to distinguish automatically comments by accused/paid trolls vs regular users.} \]

The above hypotheses test whether troll accusations can be used for distant supervision when looking for (paid) opinion manipulation trolls. However, troll accusations, trollness and troll characteristics can be also used as features for solving different (but related) tasks. Thus, we further study the relationship between “trollness” and the credibility of user’s comments, which gives rise to our final hypothesis:

\[ \text{H4. Posts by accused trolls have low credibility in the eyes of the forum users.} \]

### 4. Overview

Figure 1 shows an overview of our approach, which contains four main blocks. First, we crawl and analyze the forum structure and the user behavior in order to manually define rules that can help us detect troll accusation comments. We also identify the meta information that can be extracted, e.g. position of the comment in the thread, number of up/down votes, number of replies, etc. Second, we collect data, which include applying comment accusation rules, identifying the users that have been accused of being trolls, as well as lists of known paid opinion manipulation users and their known manipulation comments, and performing manual validation of the data. In the third step, we build a classifier from the previously annotated data in order to automatically distinguish trolls vs non-trolls (trolls could be either known paid opinion manipulation trolls or users accused by other of being trolls). This step requires defining a rich set of features, and performing feature selection on a validation set. The performance of the trained classifier is evaluated on a held out subset of the examples. Finally, the last block shows some potential applications. For instance, our classification models can provide assistance to forum moderators by pointing them to the most suspicious users and comments, thus speeding-up cleaning the forum content. Also, they can be useful to forum participants by providing
them with higher-quality information and, ultimately, better experience. Finally, the models can be used to support investigative journalists in their work by helping them to track paid opinion manipulation trolls.

5. Finding accused troll users
First, we test our first hypothesis, H1. In particular, we study whether users who have been accused of being trolls by several other forum users are special in their behavior. We further build a classifier that tries to distinguish accused trolls vs regular users. We use data from a Bulgarian news community forum.

We use leaked reports written by paid opinion manipulation trolls to their employer[1], in which they list the online forum, the user account and the comments they have posted on a particular day.

The paid troll comments from the leaked reputation management documents include 10,150 paid troll comments: 2,000 from Facebook, and 8,150 from news community forums. This gave us gold annotations for distinguishing paid opinion manipulation trolls vs regular users (Mihaylov, Koychev, Georgiev and Nakov, 2015). As the number of examples was limited, we considered as accused trolls users who were called such by at least n distinct users, and as non-trolls if they have never been called so, despite being equally active (Mihaylov, Georgiev, and Nakov, 2015). Requiring that a user should have at least 100 comments in order to be interesting for our experiments left us with 317 trolls and 964 non-trolls (if we change this to 150 comments, we have 314 accused trolls); see Table I for some statistics. We further have 15 paid trolls, who have posted in our target forum. Here are two examples of real troll comments (here, translated to English, from Bulgarian):

(1) To comment from “Historama”: Troll, you know that you cannot manipulate public opinion, right?

(2) To comment from “Rozalina”: You, trolls, are so funny I saw the same signature under other comments.

5.1 Features
To model the characteristics of “accused” trolls, we extract several features from the forum data. Our features model troll user behavior as well as observations on the behavior of users in the news community forum, targeted in this work:

- Vote-based features: we calculate the number of comments with positive and negative votes for each user. This is useful as we assume that non-trolls are likely to disagree with trolls, and to give them negative votes. We use the sum from all comments as a feature. We also count separately the comments with high, low and medium positive to negative ratio.

<table>
<thead>
<tr>
<th>Object</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publications</td>
<td>34,514</td>
</tr>
<tr>
<td>Comments</td>
<td>1,930,818</td>
</tr>
<tr>
<td>Of which replies</td>
<td>897,806</td>
</tr>
<tr>
<td>User profiles</td>
<td>14,598</td>
</tr>
<tr>
<td>Topics</td>
<td>232</td>
</tr>
<tr>
<td>Tags</td>
<td>13,575</td>
</tr>
</tbody>
</table>

Table I. Data statistics

Note: Statistics about our raw forum data from Dnevnik.bg, a Bulgarian newspaper, which we collected in April 2015, limiting the crawling to the interval from January 1, 2013 to April 1, 2015, and to three forum categories: Bulgaria, Europe and World.
• Comment-to-publication similarity: these features measure the similarity between comments and publications. We use cosine and TF-IDF[2]-weighted vectors for the comment and for the publication. The idea is that trolls might try to change or blur the topic of the publication if it differs from his/her views or agenda.

• Comment order-based features: we count how many comments the user has among the first $k$, as trolls might try to comment first to achieve higher impact.

• Top loved/hated comments: we calculate the number of times the user’s comments were among the top 1, 3, 5, 10 most loved/hated in a thread. The idea is that troll comments should be among the most hated.

• Comment reply-based features: these are features that count how many user comments are replies to other comments, how many are replies to replies, and so on. The assumption here is that trolls try to post the most comments and want to dominate the conversation, especially when defending a specific cause.

• Time-based features: we generate features from the number of comments posted during different time periods on a daily or on a weekly basis. We assume that trolls are likely to be active predominantly during specific time intervals.

Note that all the above features are scaled, i.e. divided by the number of comments, by the number of days the user has spent in the forum, by the number of days in which the user posted more than one comment, etc.

• Non-scaled features: we further compute features based on the above statistics but without scaling.

5.2 Experiments and evaluation
For each user, we normalize the features in the $[-1, 1]$ interval. We then train a support vector machine (SVM) classifier using LIBSVM (Chang and Lin, 2011), with a radial basis function (RBF) kernel, optimizing the values of $C$ and $\gamma$ empirically using grid search and five-fold cross-validation.

The results are shown in Table II. As the choice of a minimum number of accusations that we used to define a troll, namely, $n = 5$, might be seen as arbitrary, we further show results with $n = 3, 4, 5, 6$. We can see that as the number of troll accusations increases, so does the accuracy. Moreover, in all cases, our classifier outperforms the majority class baseline by a sizable margin. This shows that our definition allows us to create two groups of users (accused trolls and non-trolls) that behave differently and that can be told apart with high accuracy. Thus, we have validated $H1$.

### Table II. Classifying accused trolls vs non-trolls using different number of minimum troll accusations to define a troll (users with 100+ comments)

<table>
<thead>
<tr>
<th>Min mentions</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trolls</td>
<td>545</td>
<td>419</td>
<td>317</td>
<td>260</td>
</tr>
<tr>
<td>Non-troll</td>
<td>964</td>
<td>964</td>
<td>964</td>
<td>964</td>
</tr>
<tr>
<td>Accuracy</td>
<td>85.49</td>
<td>87.85</td>
<td>90.87</td>
<td>92.32</td>
</tr>
<tr>
<td>Improvement</td>
<td>+21.60</td>
<td>+18.15</td>
<td>+15.61</td>
<td>+13.56</td>
</tr>
</tbody>
</table>

**Note:** The Accuracy row shows the cross-validation accuracy, and the Improvement is over the majority class baseline.
vs non-troll classifier. We then apply this classifier to the task of distinguishing paid trolls vs non-trolls. Finally, we do qualitative comparison of the user profiles of paid trolls vs accused trolls vs non-trolls based on averaged feature values.

6.1 Experiments and evaluation

We only know 15 paid trolls from the leaked troll reports, which is too little to use for both training and testing. Thus, we train on accused trolls vs non-trolls classifier, but then we test on paid trolls vs non-trolls. We focus on the top four known paid trolls with the highest number of posts, as they had more than 100 comments, which means that we have enough information about them. Thus, for testing, we use the four trolls with 100 posts or more, to which we add four non-trolls (i.e. users who have never been called trolls). For training, we use exactly the same learning algorithm and the scaled features described in the previous section.

The main results are summarized in Table III. We experiment with different versions of the training data set by varying the number of accusations for a user of being a troll (by different people) needed for us to consider a user an “accused troll”; we try 3, 4, 5 and 6. This decision implies having training data sets of different sizes. In all cases, we consider users with 150+ comments and we create perfectly balanced classes (accused trolls vs non-trolls).

The second section of Table III shows the results when classifying the paid trolls from the test set. We can see that the best results are achieved when training with a minimum of four or five mentions, which yields an accuracy of 0.88. For reference, the third horizontal section in Table III shows the results when classifying the accused trolls, with cross-validation on the training data set. This should be considered as an upper bound for the classification of paid trolls, as in this case both the training and the evaluation are done with the same type of trolls (accused). The good news is that the results for paid trolls are not very far from those for accused trolls (especially when the minimum number of mentions is 4–5), indicating that training on accused trolls is a good proxy for predicting paid trolls.

Finally, we build and analyze aggregated profiles for the three kinds of users we considered: paid trolls vs accused trolls vs non-trolls. We take the average values of the most notable features for the users with the highest number of comments from each group. The values are normalized in the [0, 1] interval and displayed on a radar chart on Figure 2. Following are some interesting observations:

- Active days to all time rate (1) shows that accused trolls write at least one comment in 52 percent of their days of all time being in the forum, while non-trolls do so 36 percent of the time and paid trolls only do it 15 percent of the time. This suggests that paid trolls are less active, maybe because they only write comments when they are paid to do so.

<table>
<thead>
<tr>
<th>Minimum number of mentions</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of accused troll users</td>
<td>536</td>
<td>416</td>
<td>314</td>
<td>259</td>
</tr>
<tr>
<td>Number of non-troll users</td>
<td>536</td>
<td>416</td>
<td>314</td>
<td>259</td>
</tr>
<tr>
<td><strong>Classifying paid trolls vs non-trolls</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.75</td>
<td>0.88</td>
<td>0.88</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>Classifying accused trolls vs non-trolls (cross-validation)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>0.83</td>
<td>0.87</td>
<td>0.91</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Table III. Classification of paid trolls with 100+ posts

Note: We train on the accused trolls data set and users with 150+ comments and varying minimum number of troll accusations per user
Comments per active day (2) shows that paid trolls and accused trolls write twice as many comments per day as non-trolls.

Average comments per publication (3) shows that both paid and accused trolls post more comments per publication than non-trolls.

Downvoted by other users, strongly downvoted by other users and mediumly downvoted by other users (4, 6, and 7) show that both paid and accused trolls have much more downvoted comments than non-trolls. Yet, this is higher for paid trolls, which could mean that they have more influence compared to the self-driven accused trolls.

Strongly upvoted by other users (5) shows that accused trolls have more upvoted comments compared to paid trolls.

Reply comments rate (8) indicates that accused trolls are more likely to write comments that are replies to other users' comments compared to non-trolls, while paid trolls prefer to write specific comments and not to engage in personal “battles.” Moreover, paid trolls are more likely to write comments on working days, and during working hours, while accused trolls and non-trolls would write comments at any time, though mostly during non-working hours (work days comments rate, work time comments rate and non-work time comments rate are 9, 10 and 11, respectively).

These observations confirm our assumptions that paid trolls write comments primarily for the money, while accused trolls do so anytime, and are self-driven. Yet, note that some of our accused trolls might be actually paid.

Overall, we can conclude that $H2$ was confirmed.

7. Finding troll comments
In this section, we test $H3$. In the previous two sections, we focused on finding opinion manipulation troll users and on modeling the behavior of accused vs paid trolls. In this section, we go beyond user profile and we try to detect individual troll vs non-troll comments in a news community forum based on both text and metadata.
The ultimate goal is to be able to detect a newly-registered troll user based on his/her posts only, rather than waiting for that user to accumulate enough posts, so that either he/she can have a chance to be called a troll by several different users or we obtain enough statistics to be able to classify that user correctly using the approach described in Sections 5 and 6.

7.1 Compiling a data set for the task
We used the same forum data from Dnevnik.bg, but this time focusing on comments rather than on users. Fortunately, this time we could collect a reasonable amount of verified paid troll comments, avoiding the need for distant supervision.

Concretely, we collected the paid troll comments from the leaked reputation management documents, which included 10,150 paid troll comments: 2,000 in Facebook and 8,150 in news community forums (Mihaylov and Nakov, 2016a). The latter included 650 comments posted in the forum of dnevnik.bg, which we use in our experiments. Here is an example, translated into English:

"How about we talk with facts?! By how much have cases of lung cancer or asthma increased after the measure was introduced?! Let me tell you – by 0 percent!!! Let’s not scare people with phantoms. There are many other things that are a thousand times more harmful, but nobody forbids them."

To complement the data set with negative examples, we collected the same number of non-troll comments from users that have at least 100 comments in the forum and have never been accused of being trolls. For each paid troll comment, we selected a non-troll comment at random from the same thread in order to have similar topic and context distribution in troll vs non-troll comments.

The final data set is thus balanced, with 650 paid troll comments and 650 non-troll comments.

7.2 Modeling paid troll comments with features
We use features focusing on the textual properties of the comments:

- Bag of words: we use words and their frequencies as features.
- Bag of stems: we use words, stemmed using BulStem (Nakov, 2003a, b).
- Word n-grams: we also experiment with two- and third-word n-grams.
- Char n-grams: we further use character 3, 4-grams.
- Word prefix: for each word token, we extract the first three or four characters.
- Word suffix: for each word token, we take the last three or four characters.
- Emoticons: we extract standard HTML-based emoticons.
- Punctuation count: we count exclamation marks, dots and question marks, both single and elongated, number of words and number of all caps words.
- Metadata: we use the time of comment posting (worktime: 9:00-19:00 h vs night: 21:00-6:00 h), part of the week (work days: Monday–Friday vs weekend: Saturday–Sunday), and the rank of the comment divided by the number of comments in the thread.
- Word2Vec clusters: we trained Word2Vec on 80m words from 35k publications and 2m comments in our forum. Then, we cluster the resulting vectors, and we use these clusters as features.
- Sentiment: we used Google to translate to Bulgarian[3] the MPQA Subjectivity Lexicon (Wilson et al., 2005), the NRC Emotion Lexicon (Mohammad and Turney, 2013) and the lexicon from (Hu and Liu, 2004). Then, we reused the sentiment analysis pipeline from (Velichkov et al., 2014), which we adapted for Bulgarian.
each lexicon feature, we summed the positive/negative values for all tokens in the
text that match lexicon entries; we took the overall sum as well.

- Bad words: we use the number of bad words in the comment as a feature. We translate to Bulgarian a pre-existing list[4], which we further augment with the three most similar words for each bad word (based on the above Word2Vec model).

- Mentions: we noted that trolls used diminutive names or humiliating nicknames when referring to politicians that they did not like, but used full or family names for people that they respected. Thus, we construct lexicons with Bulgarian politician names and nicknames, and we generate a mention count feature for each lexicon.

- Part of speech (POS) tag distribution: we use GATE (Cunningham et al., 2011) to do POS tagging. Then, we use the number of occurrences of coarse- and fine-grained POS tags as features.

- Named entities: we also use the number of occurrence of named entities such as location, country, person_name as features. In future work, we plan to use better named entity recognizers, e.g. based on CRF (Georgiev et al., 2009).

7.3 Experiments and evaluation
We trained and evaluated an L2-regularized logistic regression classifier.

Table IV shows the results for telling apart comments by paid trolls vs non-trolls, when using bag of words features only, metadata only and all the above-described features together. We can see that the metadata and the bag of words features perform 22 and 25 points absolute above the baseline, respectively. The $F_1$ scores are well aligned with accuracy, and precision and recall scores are balanced. The final accuracy of 81.31 percent indicates that the classifier could be used in practice for automatic comment classification with very good overall results, which are 30 points above the majority class baseline. Thus, we can conclude that $H_3$ was confirmed.

8. Credibility in community question answering forums
In this section, we investigate $H_4$. In particular, we use our definition of trollness in order to derive features that model users of a community forum (e.g. StackOverflow, Quora, QatarLiving, etc.), and we train a system to predict the credibility of their posts[5]. Note that unlike the previous three experiments, here we do not use trollness as distant supervision for learning, but rather to obtain features for solving a different task. By conducting comparative experiments, we show that the trollness-based features are crucial to obtain the best results on credibility ranking.

8.1 Compilation of a credibility corpus
We annotated with credibility judgments question-answer threads from the SemEval-2016 Task 3 corpus that were provided as unlabeled data for the task (Nakov et al., 2016).

<table>
<thead>
<tr>
<th>Systems</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_1$</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>All features</td>
<td>83.42</td>
<td>78.00</td>
<td>80.40 (+30.40)</td>
<td>81.31 (+31.31)</td>
</tr>
<tr>
<td>Bag of words</td>
<td>76.40</td>
<td>74.31</td>
<td>75.27 (+25.27)</td>
<td>75.62 (+25.62)</td>
</tr>
<tr>
<td>Metadata</td>
<td>70.34</td>
<td>77.85</td>
<td>73.79 (+23.79)</td>
<td>72.54 (+22.54)</td>
</tr>
<tr>
<td>Baseline</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
</tr>
</tbody>
</table>

**Notes:** The absolute improvement over the baseline is shown in parentheses. The best results are in bold.

Table IV. Results for classification of comments written by paid trolls vs non-trolls.
We annotated the first ten answers in a thread with two labels: goodness and credibility. For the goodness labels, we used the task definition, assigning a Good label if “the answer or a portion of it directly answers at least one subquestion of the target question.” Regarding credibility, we define an answer as Credible if “the information in the question's thread and/or world knowledge and/or our common sense tells us that the answer is (somewhat) credible.” Table V shows some statistics about the resulting credibility data sets (Nakov, Mihaylova, Marquez, Shiroya and Koychev, 2017).

Some troll accusation in the Qatar Living forum, collected from different threads:

- Bepsy is a word used only in QL how come a new user knows this word…. You are a TROLL.
- Yes you are a troll, feeding 30 poor people cannot be an excuse, your link is totally unrelated to that….
- This is definitely a troll out to fool us. A newbie with just One point and posting a thread like this. Admin where art though when we need you.

8.2 Modeling credibility

We are interested in studying whether trollness-based features can help identify non-credible posts. We further implemented a variety of other features for comparison purposes, and also in order to create the strongest possible system.

8.2.1 Trollness features. We use features inspired by the trollness definition above, namely, that a person who is called a troll by other users is likely to be one. The list “Some troll accusation in the Qatar Living forum, collected from different threads” shows examples of troll accusations in QatarLiving. We used the number of answers that are exactly $k$ ($k = 1, 2, \ldots, 10$) answers before an answer that contains words like troll, trolls, trolling and the number of answers that come within $[0, n]$ answers before such words are mentioned ($n = 3, 5$). These 12 features have two versions: once as absolute numbers and once normalized by the total number $N$ of answers the user has posted in a thread and they were followed later by a troll mention. This total number is also a feature. We also use the average distance of the user’s post to a troll mention that comes later in the thread. Finally, we have a feature that measures the average distance not as number of answers but as average time (days). In total, this group contains 27 individual features.

8.2.2 Other features

- Quality: we train a classifier to predict whether an answer is Good. We then run it on the entire forum dump, and we aggregate its predictions to calculate the following features for each user: number of Good/Bad answers, total number of answers, percentage of Good/Bad answers, sum of the probabilities for Good/Bad answers, total sum of the probabilities over all answers, average probability for Good/Bad answer and highest probability for Good/Bad answer.
- Activity: these features describe the overall activity of the user, and are similar to the features in Section 5.1. We use number of answers posted, distinct questions to which Good answers

<table>
<thead>
<tr>
<th>Questions</th>
<th>Credible</th>
<th>Non-credible</th>
<th>Bad answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>376</td>
<td>1,733</td>
<td>73</td>
</tr>
<tr>
<td>Dev</td>
<td>50</td>
<td>137</td>
<td>15</td>
</tr>
<tr>
<td>Test</td>
<td>50</td>
<td>213</td>
<td>19</td>
</tr>
</tbody>
</table>

Table V. Statistics about our credibility data sets
an answer was posted, questions asked, posts in specific categories, days since
registering in the forum and active days. We also model the number of answers
posted in different days and hourly periods. Finally, we track the number of answers
posted among the first \( k \) in the thread.

- Credibility: we extract some features about the answer, many of which have been
  used previously for credibility (Castillo et al., 2011): number of sentences and tokens;
  of nouns, verbs, adjectives, adverbs, pronouns; of first/second/third-person pronouns;
  of URLs, images, emails, phone numbers; of positive and negative smileys; of single/
double/triple exclamation and interrogation symbols; of interrogative sentences; of
words that are not in Word2Vec’s Google News vocabulary; and, also, the average
number of tokens.

- Sentiment: we extract features modeling the sentiment polarity of the answer. We use
  NRC Hashtag Sentiment Lexicon and Sentiment140 Lexicon sentiment polarity lexicons
  from (Mohammad et al., 2013). We then used as features the sum/count/maximum of the
positive/negative values for the tokens and the bigrams in a comment.

- Goodness: we use goodness polarity lexicons, similar to sentiment lexicons, which
  were previously generated from the Qatar Living forum (Balchev et al., 2016;
Mihaylov et al., 2017).

- Google vectors: using the pre-trained 300-dimensional embedding vectors that
  Tomas Mikolov has trained on 100bn words from Google News (Mikolov et al.,
2013), we compute a vector for an answer as the average of the embedding vectors
of its words.

- QL vectors: we further use 100-dimensional word embeddings, which were
  previously trained using Word2Vec on all the available Qatar Living data (Mihaylov
and Nakov, 2016b).

- Syntactic vectors: we further use 25-dimensional embeddings of the answer
  generated by the Stanford neural parser (Socher et al., 2013), as a by-product of
syntactic parsing.

- Cosines: we calculate the cosines between an answer and the corresponding question
  using their Google, QL and syntactic vectors.

- MT features We use machine translation evaluation metrics: Bleu (Papineni et al.,
2002), NIST (Doddington, 2002), TER (Snover et al., 2006), Meteor (Lavie and
Denkowski, 2009), unigram Precision and Recall.

- BLEU components: we further use components from the computation of Bleu: \( n \)-gram
  precision, \( n \)-gram matches, number of \( n \)-grams \( (n = 1, 2, 3, 4) \), lengths of the
  hypothesis/reference, length ratio and brevity penalty (Guzmán et al., 2015, 2017).

- Rank: we further use the reciprocal rank of the answer in the thread, and percentile of
  the answer in the thread, where the first answer gets a score of 1.0, the second one
gets 0.9, the next one gets 0.8, and so on.

- Cosines thread-level: we use the cosines between an answer and a thread vector
  (all Good answers in the thread) using Google, QatarLiving and syntactic vectors.

8.3 Experiments and evaluation
We use an SVM classifier from scikit-learn, and we oversample the NonCredible to balance
the classes. We then scale the feature weights, and we use an RBF kernel, optimizing \( C \) and \( \gamma \)
on the training data set using cross-validation.
Table VI presents the results obtained by the SVM classifier when training with the five top-performing feature groups, and their combination in a metaclassifier. Since this is a ranking task (i.e. the non-credible posts in a thread should be ranked above the credible ones), we measured the quality of the output ranking with Mean Average Precision (MAP), mean reciprocal rank (MRR) and average recall (AvgRec). We further have a baseline system, which ranks the comments in chronological order, and an Oracle system, which ranks all non-credible answers above all credible ones, and therefore shows the upper bound scores. The MAP and MRR upper bounds are low as MAP and MRR are zero by definition when a thread has no non-credible answers, which is frequent in our data set.

We can see in Table VI that the best-performing group of features, on all metrics, is that containing the Trollness features. Using this feature group alone covers almost half of the gap between the MAP performance of the Chronological baseline system (13.07) and the upper bound given by the perfect Oracle (22.73), which is 44.72 percent relative MAP error reduction. This shows that users that are seen as trolls by other users in one context tend to give generally non-credible answers.

As expected, the combination of the best feature groups (Combination) improves the results of the individual groups, but the improvement over Trollness is not dramatic. It is noticeable in terms of MAP (+1.49) and AvgRec (+2.26), but for MRR the result of the combined system is not even improving over the Trollness features (−0.19). In summary, the Trollness features account for most of the performance gains over the baseline for the combined system. This confirms H4, suggesting again that our definition of trollness is indeed sensible, as trolls are hardly credible in the eyes of experienced users.

9. Discussion

In this work, we have explored the presence of opinion manipulation trolls. From a theoretical perspective, we have proposed a novel definition of what a troll is, and we have proposed a way of finding troll users and their posts, based on the assumption that a user who is called a troll by several people is likely to be one. From a practical perspective, we have shown that our definition is operational, and we have demonstrated its utility for detecting “accused” and paid opinion manipulation trolls (as a way of generating training data for distant supervision) and their comments as well as for comment credibility prediction in news community forums (as a source for features). We have further introduced and we have evaluated the utility of some novel features, e.g. “trollness,” which were not used before for credibility prediction; we have also ported some features that were used before but in a different context. Overall, our approach is general in nature, e.g. we have demonstrated the portability of our trollness definition as well as of some of the features across tasks, domains, forums and languages.

<table>
<thead>
<tr>
<th>System</th>
<th>MAP</th>
<th>AvgRec</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Trollness</td>
<td>17.39</td>
<td>91.19</td>
<td>19.51</td>
</tr>
<tr>
<td>2. Syntactic vectors</td>
<td>16.57</td>
<td>88.14</td>
<td>17.42</td>
</tr>
<tr>
<td>3. Cosines</td>
<td>16.31</td>
<td>84.78</td>
<td>16.54</td>
</tr>
<tr>
<td>4. Quality</td>
<td>15.98</td>
<td>90.24</td>
<td>17.05</td>
</tr>
<tr>
<td>5. Credibility</td>
<td>15.38</td>
<td>89.29</td>
<td>15.57</td>
</tr>
<tr>
<td>Combination</td>
<td>18.88</td>
<td>93.45</td>
<td>19.32</td>
</tr>
<tr>
<td>Chronological</td>
<td>13.07</td>
<td>83.84</td>
<td>13.14</td>
</tr>
<tr>
<td>Oracle</td>
<td>22.73</td>
<td>100.00</td>
<td>22.73</td>
</tr>
</tbody>
</table>

Notes: We show the performance for the best five feature groups and for their combination, as well as for the baseline and an oracle upper bound. The best scores with individual features are in bold. The values in italic indicate Baseline and Oracle scores.

Table VI. Evaluation results
9.1 Theoretical implications
Based on previous observations in the literature that labeling a forum user as a troll is a common way of accusing them of improper behavior (Gazan, 2016), we have proposed a novel definition of what a troll is, based on the assumption that a user who is labeled a troll by several forum users is likely to be one. We have further demonstrated the utility of this definition for detecting accused and paid opinion manipulation trolls and their comments as well as for predicting the credibility of comments in news community forums. We believe that these results open an interesting direction for future research, by empowering researchers with a useful, practical definition of trollness.

More generally, we have shown that troll users are easily identifiable, despite previous observations in the literature that trolls typically abide by the formal forum rules and try to leave an impression of being good members of the forum community (Donath, 1999).

Moreover, this is the first study on whether forum users accused of being trolls really behave like such (e.g. known paid trolls) and also whether they behave differently from users who have never been labeled as trolls. In particular, we studied the nature of the potential differences in behavior, pointing to important differences between accused vs paid trolls vs non-trolls. This enables further research in understanding the differences between the behavior of trolls and normal forum users, which again would be facilitated by the practicality of our definition for troll users.

Finally, we have demonstrated for the first time that there is correlation between trollness and credibility of comments in a news community forum. This opens yet another possible new research direction into the relationship between veracity of information and trollness.

9.2 Practical implications
Our findings suggest that the proposed approach would help forum moderators in their work, e.g., by pointing them to the most suspicious users and comments. It might be also useful to investigative journalists who want to track the paid opinion manipulation trolls. Managers running news portals with community forums should realize that that they can offer their users improved experience by identifying and filtering out paid opinion manipulation trolls; this would also boost the credibility and the reputation of their online community forum as well as of their news portal in general.

9.3 Limitations
Below, we present a critical discussion of the described research, which we organize around three topics: reliability, application and transferability.

9.3.1 Reliability. We have considered two kinds of opinion manipulation trolls. First, we studied paid trolls, who were hired by a political entity to write comments in internet forums and social media, and who were given specific key points to defend on a daily basis. We further defined the notion of accused trolls, who were called trolls by several other users in the forum.

We believe that the annotations about which users in the forum are paid trolls is quite reliable, as we have obtained the list from a set of leaked documents, which include copies of contracts between a company and people hired as paid opinion manipulation trolls, as well as reports from these paid trolls about the actual work that they have done as an obligation to their contracts.

However, we are less confident in the reliability of the annotations for the accused trolls. By our definition, these users were accused of being trolls by several other users in the forum. The problem is that we have no control over who these accusers are, and we do not know their motivation to do the accusation. It could be that “real” trolls could have accused non-trolls, e.g. as an argument in dispute. Yet, we have shown, both experimentally and analytically, that the accused trolls have quite different user-level characteristics and behavior compared to non-trolls, and that they are similar to paid trolls in many aspects.
Finally, we have transferred our trollness definition to a different domain, namely, community question answering, and we have shown that it helps to identify non-credible answers in the forum. This is an indirect indication that our definition is reasonable and that it identifies a group of users who are not trusted by other users in the forum.

Our study on identifying trolls and on credibility detection depends on users having to sign into the system. However, a user can easily create multiple identities (i.e. sockpuppeting), in order to simulate discussion, sympathy, respect and credibility. A related issue that may limit the technical applicability of our approach is that some forums do not require a user to sign in to be able to post.

9.3.2 Application. We have shown that we can accurately detect trolls in a news community forum, and that we can use trollness-based features in order to improve credibility detection in another community question answering forum. All this was done using static data sets. However, incorporating our model in an online, robust, and adaptive application would present some further challenges, e.g. we would need to be able to take new users, comments, replies, likes, etc., into account as soon as possible after they have come into existence.

This is especially important when hunting for troll comments. While user-related features, such as consistence, frequency and time of posting, might remain generally stable over time, paid troll comments are much more dynamic in nature, as the people posting them are often instructed to support a given list of key points on concrete topics that change daily. Thus, while training on comments on a particular topic and time might generalize to close periods of time when the topic is still hot, the performance would inevitably degrade over time.

9.3.3 Technical transferability. We believe that our approach is generally transferable to other forums, domains and languages. In fact, our experiments above have already shown that the notion of trollness is useful across four tasks, two forums and two languages.

First, our definition of accused troll only implies to check whether a certain username appears in the context of the word troll and all its variants. This can easily be transferred to other communities by counting the number of mentions of a given username in the context of words like troll, trolls and the like. Thus, transfer across languages only requires minimal adaptation.

Second, we have defined a set of features based on the behavior of internet trolls, e.g. frequency of posting comments and replies, time of posting, etc., which allow us to expose troll comments. These features can be extracted and used for most news community forums, which typically have a similar structure.

Regarding the detection of troll posts, we also use the post content. In this case, our approach is language-dependent, as we developed it for Bulgarian. However, most of the features that we have defined can be transferred to other languages; all that would be needed are standard tools such as POS taggers and named entity recognizers. In order to extract the sentiment features, we automatically translated popular sentiment lexicons from other languages, which is another way of adapting resources. In summary, language adaptation should be lightweight in terms of preprocessing and feature extraction.

9.4 Future work

In future work, we want to try our trollness definition, and some possible variations thereof, on other tasks, domains, and languages. We further plan to explore the use of neural networks, e.g. using long short-term memory or convolutional neural networks to obtain a distributed representation for a comment.

We could then combine such representations in complex neural network architectures, which we could possibly train using additional data in a semi-supervised manner, with distant supervision. We would also like to analyze comment threads as a whole (Barrón-Cedeño et al., 2015; Joty et al., 2015, 2016), and to try to learn from the set of
comments that have been deleted by the moderator of the forum, as well as from user behavior across forum boundaries, e.g. we know that some of our paid trolls were active on several forums. Detecting user cloning on the same site or across websites, i.e., sockpuppets, or identifying coordinated opinion manipulation by several users who promote a similar position around the same time, i.e. internet water army, is another research direction that we see important. Overall, we want to contribute to the best of our knowledge and ability to the never-ending fight against the dark side of the Web.

Notes

2. The term frequency-inverse document frequency (TF-IDF) is a measure that shows how important a word is for a document in a given collection of documents.
3. In future work, we plan to add sentiment analysis lexicons developed for Bulgarian (Kapukaranov and Nakov, 2015), or for the closely related Macedonian (Jovanoski et al., 2015, 2016).
5. This aspect is ignored, e.g. in recent Community question answering tasks at SemEval (Nakov et al., 2015, 2016; Nakov, Hoogeveen, Márquez, Moschitti, Mubarak, Baldwi and Verspoor, 2017) where an answer is considered Good if it tries to answer the question, irrespective of its veracity.

References


The dark side of news community forums


Corresponding author
Todor Mihaylov can be contacted at: tbmihailov@gmail.com

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Online moral disengagement and hostile emotions in discussions on hosting immigrants

Francesca D’Errico

Department of Philosophy, Communication and Visual Arts,
University of Roma Tre, Rome, Italy, and
Marinella Paciello

Faculty of Psychology, Uninettuno University, Rome, Italy

Abstract

Purpose – The purpose of this paper is to explore the “dark nuances” of social media by identifying moral disengagement (MD) mechanisms and hostile emotions in people discussing the hosting of immigrants and examining the relationship between MD mechanisms and hostile emotions expressed online (annoyance, irritation and contempt).

Design/methodology/approach – The method was based on a psycho-lexicographical approach that analysed the in vivo real words, adopting a quanti-qualitative point of view. The investigation started from the case of a Facebook post in support of immigrants after a serious shipwreck causing the death of more than 700 would-be migrants. More than 10,000 comments were codified. For the comments against the hosting of such immigrants, the authors followed a codebook aimed at identifying MD mechanisms and hostile emotions.

Findings – The main findings show an interplay between different hostile negative emotions and online MD mechanisms. The greater the intensity of hostile emotions, the more the locus of disengagement moves from the unethical individual’s behaviour – for example, offering moral justifications – to the target recipient of such behaviour – for example, blaming or dehumanising.

Practical implications – The study could be applicable in designing and developing algorithms and technological tools aimed at preventive interventions to promote moral awareness and emotional regulation in online settings.

Social implications – The study may be a useful contribution to understanding unethical orientation by identifying areas where education can intervene in reducing harmful behavioural tendencies.

Originality/value – This study takes together expressed hostile emotions and MD mechanisms by means of the analysis of real words in vivo through social media discussions.

Keywords Social media, Moral disengagement, Hostile emotions, Hosting immigrants, Online discussions

Paper type Research paper

1. Introduction

In recent years, the numbers of international migrants to Europe and North America have greatly increased. At the same time, immigration has come to be perceived, alongside terrorism, as one of the most threatening social risks in terms of local identities and security (Karamanidou, 2015). In this framework, classical and new media communications have played a crucial role, feeding a negative image of immigrants by strengthening the perception of social threat and polarising their representation as desperate individuals or criminals who cross national borders illegally (Esses et al., 2013; Bruno, 2016). If the popular media tend to frame immigrants as an embarrassment or a danger, by contrast, international institutions such as the United Nations and in particular its refugee agency (UNHCR) have signed several conventions acknowledging the necessity of a “moral standard and promotion of immigrants” rights across nations, by governing the protection of migrants and their families (UNHCR, 2010, 2011).

Taking these considerations into account, this paper examines social media discussions on the hosting of immigrants and the opportunity for groups and institutions to offer asylum, help or a place of safety to immigrants – people who, for affective or personal...
reasons or because of political persecution and war, move to a country they were not born in. Within this debate, people may simply agree or disagree about accepting immigrants without arguing their position, or they may present ethical or unethical arguments for helping or temporarily accepting them in their countries.

The aim of the present work is to explore the context of social media discussions on the hosting of immigrants by considering the moral disengagement (MD) mechanisms that may underlie an individual’s potentially negative stance towards accepting immigrants, going beyond the simple polarisation between racists and non-racists. To this end, the paper refers to the theoretical field of moral agency (Bandura, 2015), representing the capability to adhere to acknowledged standards, and focussing, in the present case, on the online ethical orientation of accepting and helping immigrants, viewed as a form of adherence to the specific moral domain of “care” (Haidt, 2001).

Numerous researchers have focussed only on the so-called “dark side” of social media by pointing out individual and contextual factors that increase harmful online behaviours (Siegel et al., 1986; Hutchens et al., 2015; D’Errico and Poggi, 2014; Poggi et al., 2015; Kwon and Gruzd, 2017), such as explicit forms of aggression (e.g. flaming and cyber-bullying), racism and discrimination. These studies have neglected the “moral perspective” that can interpret the stance of being in favour of hosting immigrants as an act of potential assistance to people in need and, on the other hand, the stance of being against hosting immigrants as a potential moral omission. Moreover, when people avoid supporting someone in need, they may be in a negative affective state and able to give more or less morally acceptable explanations for their lack of helpfulness.

The aim of the present study is to examine the interplay of hostile emotions and morally disengaged arguments emerging from the words used in social media discussions, specifically by investigating the case of a Facebook post in favour of hosting immigrants promoted by a public figure (Gianni Morandi, a popular Italian singer). In the “Morandi case”, a public communicator, following a serious accident at sea that caused the death of more than 700 people, tried to promote a positive and ethical orientation among his followers towards hosting immigrants, by means of a public message on his Facebook page. This post sparked a popular debate, in which people publicly commented and stated where they stood in relation to the hosting of immigrants, sometimes in aggressive and hostile tones. In particular, the present study investigates, first, the different moral arguments given by users to justify their potential “omission”, that is, their lack of willingness to provide help to immigrants. Omitting to help, unlike direct aggression, is a “passive” negative behaviour related to the universal moral domain of care/harm (Graham et al., 2011; Haidt and Joseph, 2007), yet it is potentially just as damaging for those in need as “active” harmful behaviour. Thus, it is possible that the same cognitive processes usually implicated in explaining aggressive conduct are used by people when legitimising an unethical and potentially damaging orientation towards others in need – in this case, immigrants.

At the same time, this study considers the role of hostile emotions usually associated with harmful behaviours (Izard, 1975; Anderson and Bushman, 2002), which can also be readily expressed and spread in computer-mediated communication (CMC), thereby promoting a negative “emotional contagion” (Kramer et al., 2014) more easily than in face-to-face interactions (Derks et al., 2008).

The relationship between MD mechanisms and negative emotions has not been explored in the online setting, although the literature on morality (Haidt, 2001) has underlined the importance of jointly studying the cognitive and affective dimensions involved in orienting moral decisions. However, the aim of the present study is not just to confirm the possible association between MD and negative emotions in online settings but also to examine in depth this relationship in terms of all the MD mechanisms theorised by Bandura (1991) and the kind of hostile emotions that are strongly related to harmful
behaviour (Anderson and Bushman, 2002; Poggi et al., 2015). It is postulated that the interplay between hostile emotions, which at a certain level can be more difficult to regulate (Gross, 2013), and MD mechanisms, which reflect the different ways in which an individual can legitimate his or her own unethical behaviour, can constitute particular unethical orientations. In particular, it is hypothesised that more intense hostile emotions are associated with MD mechanisms that increase the distance between the potential helper and others in need (i.e. immigrants), such as blame attribution and dehumanisation. As to other MD mechanisms, this issue has never been explored in such an analytical way, either in the case of passive orientations and behaviours (i.e. omission) or in the case of active forms (i.e. verbal aggression and flaming).

2. Moral disengagement
The construct of MD was developed within Bandura’s theory of moral agency (Bandura, 1991, 2015), one of the most influential approaches in the morality-behaviour arena. According to this theoretical framework, people are active agents and through self-regulatory processes are able to control their moral behaviour, namely, to align moral conduct with moral and ethical principles. More specifically, the moral self-regulative process operates by means of functions of human agency, such as self-monitoring, self-evaluation and self-reflection, and it can take a proactive or an inhibitive form. The proactive process fosters moral action by adjusting it according to personal and social standards and by anticipating positive emotional self-evaluative reactions. For example, when people believe that it is important to promote respect for human rights, they may be proactively engaged in a plan of action aimed at protecting and defending this principle and, in so doing, they anticipate positive emotions such as pride and satisfaction, derived from the possible advantageous consequences of their choices and actions. The inhibitive process instead hinders engagement in immoral behaviours by construing them as ethically and socially sanctionable and by anticipating negative emotional self-evaluative reactions. For example, in the case of human rights, people may refrain from action because this principle does not accord with their own, and so they avoid the possible negative emotions such as guilt and shame that may result from the negative effects of their unethical action (Bandura, 1991). In other words, moral self-regulation serves to promote ethical conduct, such as helping and caring for others in need, and to prevent unethical behaviour, such as harmful conduct, allowing people to remain anchored to the norms and principles that guide their moral behaviour.

However, the moral self-regulation process does not always work in a consistent and constant way (Bandura, 1991, 2015). Under certain circumstances, some social cognitive processes, defined as MD mechanisms, can “darken” moral purpose and deactivate or diminish moral control. MD mechanisms are cognitive stratagems that allow people to avoid self-blame and protect their moral image by distorting or refraining from certain behaviours and their consequences and by locating the responsibility for their actions outside themselves. Through MD mechanisms, “good” people can also temporarily silence their conscience and so exhibit behaviour they usually condemn in other situations. There are eight MD mechanisms identified by Bandura (1991), and these operate at different levels (loci): at the behaviour locus, moral justification, euphemistic labelling and advantageous comparison redefine the behaviour itself; at the outcome locus, the disregarding/distorting consequences alter the perception of the effects of behaviour; at the agency locus, displacement and diffusion of responsibility obscure personal accountability, moving it from the actor to others or to situational circumstances; finally, at the recipient locus, dehumanisation and attribution of blame operate, respectively, to divest people of human qualities and to attribute to them the fault of injurious conduct that becomes a justifiable defensive reaction to instigations.
All these mechanisms have been tracked in studies that have focussed on understanding why individuals behave aggressively and unethically. A large body of research has demonstrated the disinhibitory power of MD in fostering harmful and rule-breaking behaviour across different age groups (Bandura et al., 1996; Caprara et al., 2008; Paciello et al., 2008; Gini et al., 2014), in different social and cultural contexts (Aquino et al., 2007; Grussendorf et al., 2002) and in clinical (Hyde et al., 2010; Muratori et al., 2016), educational (Pozzoli et al., 2012; Thornberg and Jungert, 2014), organisational (Fida et al., 2015; Moore, 2008) and sports settings (Proios, 2016). With regard to the specific case of omission, only a few studies on adolescents have investigated the role of the MD mechanism in explaining why individuals do not behave in a prosocial and moral way (Paciello et al., 2013), the assumption being also that these mechanisms take the same form as in the case of harmful and aggressive behaviour.

3. MD and online unethical behaviours

Recently, the disinhibitory role of MD has been confirmed also in the case of online unethical behaviour, particularly behaviour that is explicitly aggressive. Indeed, attention has been focussed mainly on the clear forms of online deviant and aggressive manifestations, such as cyber-aggression and cybercrime (Lowry et al., 2017; Rogers, 2003). Most contributions on the role of MD derive from the growing literature on cyber-bullying (Bauman, 2010; Cross et al., 2015; Perren and Gutzwiller-Helfenfinger, 2012; Pornari and Wood, 2010). Studies focussed on children and adolescents have suggested that MD is related to both traditional bullying and cyber-bullying (Bussey et al., 2015; Chen et al., 2016; Wang et al., 2016). However, certain results are still controversial. Some authors suggest that MD is strongly related to traditional bullying only and that, when it comes to adolescents’ online aggressive behaviours, this relationship is inconsistent (Bauman and Pero, 2010; Perren and Gutzwiller-Helfenfinger, 2012; Pornari and Wood, 2010). In particular, Pornari and Wood (2010) suggest that the contribution of MD to cyber-bullying is weaker than its contribution to traditional bullying, because some characteristics of the online environment (anonymity and distance from the victim) make adolescents less aware of the possible consequences of their actions and less prone to feel negative emotions. By contrast, other researchers have suggested that, in online settings, some technological affordances could increase the likelihood to disengage morally. The characteristics of the virtual environment that seem to increase accessibility to MD mechanisms are the paucity of social emotional cues, the ease of disseminating communication via social networks, increased media attention to online aggressive forms and the lack of space and time constraints on aggressive manifestations (Runions and Bak, 2015). Moreover, some authors have attributed the anonymity and lack of face-to-face contact to an increase in MD in technologically mediated contexts (Cross et al., 2015). Thus, an alternative hypothesis seems to be that CMC could facilitate recourse to MD mechanisms, given the greater perceived distance between self and victim, which decreases empathic contact. Since empathy is one of the most important protective factors in the prevention of MD (Paciello et al., 2013; Detert et al., 2008), this affective distance could facilitate engagement in harmful behaviour by activating self-serving and self-focussed cognitive processes, such as MD mechanisms.

With the exception of online cyber-bullying, few researchers have examined the relationship between MD and other online unethical behaviours, but the literature is continuously increasing. In particular, two empirical studies have shown the association between MD and explicit and aggressive unethical expressions, such as online racism (Faulkner and Bliuc, 2016) and violent ideological websites (Connelly et al., 2016). Furthermore, Faulkner and Bliuc (2016) highlighted the relevance of MD in understanding online responses to racist incidents. About 90 per cent of online racist comments use one or more MD strategies, such as euphemistic labelling and advantageous comparison for
reframing racist action, displacement of responsibility or assigning blame to the victims. By contrast, none of the MD mechanisms has been identified in comments against forms of racism. Connelly et al. (2016) found that ideological groups’ websites include content reflecting dehumanisation, misrepresentation of consequences, euphemistic labelling and displacement and diffusion of responsibility. These contributions have partially mapped MD mechanisms, since they started from aggressive and extreme comments, by necessarily neglecting “unsupportive” comments where people just expressed their lack of willingness to help or host immigrants.

To the present authors’ knowledge, only these two empirical studies have focused on this topic and, as the researchers themselves suggest, it could be useful to explore in depth how MD can take form and be spread in online discussions. Moreover, examination in vivo of MD mechanisms in the online setting could be particularly important, since MD may play a key role here in legitimising forms of collective violence (Leidner et al., 2010). Indeed, as theorised by Bandura, MD has not only a cognitive but also a social nature, and different MD strategies “usually operate in concert rather than in isolation at both the individual and social systems level” (Bandura, 2007, p. 11). Thus, it is plausible that, in online interactions, in which people share beliefs, modes of communication and rules, the MD mechanisms could be learned, influencing ethical orientation. From this perspective, social networks represent a “social square”, in which people from broad geographical areas can meet virtually and reinforce their cognitive processes, thereby making unethical ideas such as extremist ideologies more acceptable (Gerstenfeld et al., 2003). In summary, the online context in which people interact and socialise may be a context that promotes MD.

4. Emotions and MD

In social networks, negative emotions such as contempt and disgust can be easily expressed and spread (Kramer et al., 2014), potentially prompting an exponential growth of racism, flaming (Siegel et al., 1986) and dehumanisation, much as in “real” interactions (Kelman, 1976; Bar-Tal, 2000; Haslam, 2006). The present study suggests that emotions can signal an activation because they represent adaptive internal devices that monitor the state of the achievement or the thwarting of one’s personal goals (Castelfranchi and Miceli, 2009; D’Errico and Poggi, 2016). In particular, so-called “hostile emotions” essentially entail a direct or indirect attribution to another person of negative or blameworthy features (Izard, 1975; Ortony et al., 1990). Hostile emotions such as distress, frustration, anger, contempt, disgust and hate potentially promote harmful and aggressive behaviours (Berkowitz, 1989; Zillmann, 1988) because they feature medium/high arousal that, in the presence of other aversive stimuli, can progressively increase. This phenomenon is the essence of the frustration-aggression hypothesis (F–A model) (Dollard et al., 1939) and its extension (Berkowitz, 1989). In particular, hostile emotions, such as anger, could direct attention to provoking events and a hostile interpretation of neutral situations (Dodge and Coie, 1987). People who are negatively activated by events perceived as frustrating and unfair tend to infer other intentions, based on anger-related knowledge structures that promote hostile cognitive processes and aggressive behaviour (Anderson and Bushman, 2002).

Whereas hostile emotions indicate how people feel about moral issues (affective morality), MD instead provides information about how individuals think about rules of ethical conduct (cognitive morality) (Koops et al., 2010). Traditionally, only moral cognitive processes have been addressed within the extensive psychological literature on cognitive developmental frameworks (Kohlberg, 1958, 1971; Rest, 1979; Turiel, 1983), and only recently has it been recognised that cognitive dimensions need to be jointly examined with emotional dimensions in order to understand moral functioning (Greene et al., 2001; Haidt, 2001). Moreover, some scholars suggest that the specific emotions involved in the moral process can be identified. For example, in relation to the harm/care moral domain,
Rozin et al. (1999) emphasised the role of anger, disgust or contempt in guiding moral decision making.

Recent findings in the specific literature on emotions and MD, focussed on “positive” emotions in relation to harmful behaviours, suggest that while enjoyment in virtual contexts can be positively associated with MD (Hartmann and Vorderer, 2010), empathy tends to hinder the activation of MD (Menesini et al., 2003; Paciello et al., 2013). The results are more complex in the case of negative emotions. Indeed, for some authors, a lack of guilt is associated with MD (Muratori et al., 2016), whereas for other scholars (Johnson and Connelly, 2016), guilt operates like a moderator that attenuates the negative relationship between MD and ethical decision making. Some scholars have found hostile emotions to be directly associated with MD (Rubio-Garay et al., 2016), whereas others believe the link is indirect. Indeed, in the longitudinal study of Caprara et al. (2014), the evidence suggests that irritability (as an individual tendency) is indirectly related to MD through hostile rumination. Thus, anger could lead to MD only when it persists over time and is converted into vengeful resentment. In addition, some recent studies focussed on contextual aspects have suggested that negative emotional activation resulting from situations perceived as stressful and frustrating can facilitate recourse to MD (Fida et al., 2015).

Overall, MD has been measured in all these studies using self-reporting methods, in which MD is considered as a one-dimensional construct (a sort of general attitude), disregarding the different loci of MD – behaviours, agency, outcomes and recipients. Moreover, the literature has never examined the relationship between MD and different levels of emotional activation that could make people less able to exercise moral control. Detecting MD mechanisms and hostile emotions within online interactions, by means of verbal expressions, can represent an in vivo observation of these two theoretical constructs by also increasing ecological validity; the online setting can contribute to providing more contextualised hints to a research topic that has been receiving increasing attention over the last few years. Indeed, previous studies have mainly focussed only on cognitive processes related to aggressive forms of verbal expression (Connelly et al., 2016; Faulkner and Bliuc, 2016). By contrast, the present study aims not only to study distinct MD mechanisms enlarging the perspective to “unsupportive comments” about immigrants, but also to examine the relationships between MD mechanisms and different hostile emotional states expressed through the words, in order to capture these possible relationships.

5. Present study
The present study was based on one case of communication from Gianni Morandi, a popular Italian singer, who posted on his personal Facebook page a message aimed at promoting closeness between Italians and would-be immigrants who tried to land on the Italian coast. The “Morandi case” can be considered a case of a public communication regarding a serious accident involving immigrants, aimed at promoting a positive and ethical orientation of the communicator’s followers towards the hosting of immigrants. Morandi’s communicative act caused a great polarisation between those who supported and those who attacked his position. The present study focusses on the different nuances of the “dark side” – that is, on unsupportive comments towards hosting immigrants.

The goals of this study were as follows: the identification of expressed hostile emotions (Izard, 1975) and MD mechanisms (Bandura, 1991, 2015) in vivo through the coding of his followers’ comments to Morandi’s supportive message; and the analysis of the relationship between hostile emotions and MD mechanisms. With regard to the first research goal, it was expected that different hostile emotions and a wide range of MD mechanisms would be detected, since the comments in reply to Morandi’s message did not come from a simple politically polarised audience, but from a very large audience. So, also in line with the theory of moral agency (Bandura, 1991), people can use a wide range of refined strategies to defend
their own moral image when they need to justify their unwillingness to help others in need. With regard to the second research goal, it was expected that a strong relationship between negative expressed emotions and MD would be found; specifically, it was hypothesised that individuals who express more hostile emotions will argue for and justify their potential failure to offer help to immigrants, by using more direct mechanisms such as attributing blame to the victims and dehumanising them.

5.1 Case selection and online communicative scenario
The communication by Gianni Morandi on 21 April 2015 is follows: the post was aimed at fostering closeness by means of social media (Facebook) between Italians and would-be immigrants after a serious shipwreck in a channel near Sicily on the night of the 18th and 19th of April, causing the death of more than 700 migrants (Kingsley et al., 2015). The message was composed of text (see below) and an image. Both text and image were intended to generate feelings of closeness and affinity by comparing the Italian emigration of the early twentieth century with the plight of the immigrants coming from Africa in recent years:

About migrants and emigrants, we must never forget that thousands and thousands of Italians, in the last century, have left their homeland to America, Germany, Australia, Canada […] with the hope of finding work, a better future for their children, because in their country they could not get it with the humiliation, harassment, abuse of power and violence, who have had to endure! It was not that long […]

This post was selected for three main reasons:

(1) Gianni Morandi is not a politician but a popular national singer, followed by people of different generations (from the 1960s to the present). Several politicians posted messages on Facebook in the same period (e.g. Matteo Salvini for the right-wing party, Lega Nord), but these were not chosen because they were politically biased.

(2) The Morandi social media message became a “social media case”, because it provoked a very large number of comments (not only from Gianni Morandi’s followers), as highlighted also by the national and international press. Furthermore, Morandi used only his personal Facebook page, not other media, such as television or newspapers.

(3) The nature of the message supporting the hosting of immigrants generated comments appropriate for studying, in terms of the harm/care moral domain (Graham et al., 2011), as well as the potential unethical orientation (i.e. omission of assistance to persons in need) of social media users[1].

5.2 Data and procedure
A total of 12,583 comments made in response to Gianni Morandi’s post, distributed between 21 and 27 April 2015, were extracted through Facebook API on 27 April 2015. For encoding, comments directed at Morandi, both positive and negative, a simple expression of agreement or disagreement (Poggi et al., 2011) and more than 400 comments containing links to videos or images were filtered. In addition to this coding, comments that simply denied the potential comparison between the two emigrations (e.g. “We are not like them” does not include either ethical or unethical argumentation) or that merely celebrated responders’ own national identity were excluded, corresponding to about 2,446 posts. The robustness of the filtering process was tested by Cohen’s $\kappa$ coefficient on “codifiable” comments (in terms of discrimination, moral reasoning, prosocial comments and personal vents) vs “non-codifiable comments” (simple agreement or disagreement on the topic, simple
denial of comparison with other groups or insults and expressions of gratitude directed towards Morandi). From the starting database, 500 comments were selected at random by reaching a very good agreement among judges (κ Cohen = 0.96; 97.30 per cent of agreements).

Data cleaning carried out in this way led to the identification of four main categories: discrimination, MD mechanisms, prosocial comments and personal vents.

As a first step, explicit discrimination was coded when the comment was directly aimed at causing harm (e.g. “Throw them in the sea!”). The prosocial comments (D’Errico, 2016) featured arguments in support of hosting immigrants (e.g. “We are all humans!”). A generic personal vent was a comment that contained a description of a negative personal situation (e.g. “My grandparents were checked at the border and had no health care”).

Thereafter, three expert judges (one MD expert, one hostile emotions expert and a naïve judge) shared theoretical definitions (Table I) and then they coded comments for MD mechanisms and also hostile emotions. All judges discussed the coding criteria starting from the subset of 500 comments and fixed a common coding grid. Subsequently, on the basis of the grid, they separately encoded the same comments (n = 1,000) in order to measure the agreements that could be considered “good” for both MD mechanisms (κ Cohen = 0.78; 85.39 per cent of agreements) and hostile emotions (κ Cohen = 0.72; 81.20 per cent of agreements). All remaining comments were randomly assigned and separately coded.

<table>
<thead>
<tr>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Locus behaviour</strong></td>
<td></td>
</tr>
<tr>
<td>Advantageous comparison</td>
<td>Unethical conduct can lose its repugnancy by comparing it with more flagrant inhumanities</td>
</tr>
<tr>
<td>Euphemistic labelling</td>
<td>The use of sanitizing and convoluted language permit to mask unethical behaviour conferring it a respectable status</td>
</tr>
<tr>
<td>Moral justification</td>
<td>By using moral explanations to challenge rational decision making norms, unethical behaviours are regarded as serving worthy purposes</td>
</tr>
<tr>
<td><strong>Locus outcome</strong></td>
<td></td>
</tr>
<tr>
<td>Consequences distortion (of hosting)</td>
<td>People minimise or distort the consequences of acts they are responsible for. Any evidence of harm by victims are discredited</td>
</tr>
<tr>
<td><strong>Locus Agency</strong></td>
<td></td>
</tr>
<tr>
<td>Displacement of responsibility</td>
<td>Individuals avoid the responsibility for their actions by viewing unethical behaviour as arising from social conditions and from authorities</td>
</tr>
<tr>
<td>Diffusion of responsibility</td>
<td>Personal accountability for one’s contribution to unethical behaviour is reduced by giving responsibility to the others and various facets of decision making to sub-groups</td>
</tr>
<tr>
<td><strong>Locus recipient</strong></td>
<td></td>
</tr>
<tr>
<td>Blame attribution</td>
<td>People view themselves as faultless victims driven to injurious conduct by forcible provocation. Unethical conduct thus becomes a justifiable defensive reaction to instigations</td>
</tr>
<tr>
<td>Dehumanisation</td>
<td>People view the victim as subhuman, and not as a person with feelings</td>
</tr>
</tbody>
</table>

Table I. Codebook of moral disengagement
With regard to MD mechanisms, Bandura’s theory and the previous codebook used in the literature (Bandura, 2006; White et al., 2009) (see Table I) were used to code online comments. In particular, each comment was analysed semantically in order to verify if and how, in arguing their positions, the participants used one of the eight moral mechanisms. In greater detail, the comments characterised by semantically restructuring the meaning of omission were classified into one mechanism operating at a behavioural level (moral justification, euphemistic language, advantageous comparison) on the basis of linguistic operations (i.e. justification, redefinition and comparison). The comments characterised by reasoning aimed at obscuring personal responsibility were classified into one mechanism operating at an agency level (dislocation and diffusion of responsibility) based on the external factor emerging from the comments (i.e. authority and socio-demographic conditions vs other bystanders). The comments characterised by a distorted interpretation of omission’s effects were classified in the mechanism operating at an outcome level (distortion of consequences). Finally, the comments characterised by reasoning aimed at attributing to the victim the fault for an unsupportive stance were classified into one mechanism operating at a target level (dehumanisation and attribution of blame), based on the way in which the potential targets were viewed (i.e. lacking human qualities or with aggressive behavioural intentions).

The analysis of the emotional expression of the work was based on psychological and psycho-linguistic models that detect emotions through both lexical and semantic levels (Ortony and Clore, 1981; Pennebaker et al., 2003; Poggi, 2001; Novielli et al., 2015) and paralinguistic aspects of their expression (Sproull and Kiesler, 1986). Specifically, analysis referred to the area of “hostile emotions” (Izard, 1975), “reproach emotions” (Ortony et al., 1990) or “aggressivity emotions” (Poggi, 2008). As to the paralinguistic side, the coding of the emotional expression was referred to as mind markers, all those signals that provide information on the speaker’s mental state, namely, his or her beliefs, goals and emotions (Poggi, 2001).

The psycho-linguistic approach deeply analysed the possible markers that in written language can be expressed grammatically (i.e. by means of exclamative sentences and emphatic syntactic structures), lexically (by emotion nouns, verbs or adjectives, such as envy, fear, angry, scary) or paralinguistically (i.e. with interjections such as wow!, in Italian ihol! = I’m surprised; Poggi, 2001). Also, Caffi and Janney (1994) identified some different “emotive devices” based on classical dimensions of affect, named evaluation and potency, by clustering not only the semantic import but also the rhetorical, prosodic and paralinguistic markers. The so-called “affective tonality” considered here included evaluative polarity (positive vs negative stance, i.e. “I am disgusted when I see immigrants around me” / “I am happy to see them in a safe place”), specificity polarity (clear or vague stance, i.e. “Immigration is very dangerous” / “Immigration, in some cases, can be a problem”), evidentiality (confident vs doubtful, i.e. “It is DANGEROUS for Italians!” vs “It might be dangerous for public safety”), volitional polarity (assertive/non-assertive, i.e. “I WANT them out of our nation” / “Do you ever think of leaving them in their own countries?”) and quantity (more or less intense, i.e. “IT IS DEFINITELY DANGEROUS for ITALIANS!!!!!” vs “It is dangerous for us”).

With regard to the hostile expression of emotions (and thus not internally felt), direct or indirect attribution to another person of negative or blameworthy features was considered (D’Errico and Poggi, 2012), which, in the case of contempt, for example, can be played by means of irony, sarcasm or insults (D’Errico and Poggi, 2014, 2016; D’Errico et al., 2014; Poggi and D’Errico, 2009; Poggi et al., 2013, 2015; Kwon and Gruzd, 2017). The hostile emotions included for the present study (linguistic and paralinguistic) were markers that expressed, respectively, annoyance, irritation and contempt (see Table II), considered as a continuum of hostility gradually more intense and focussed against immigrants.
Given the present descriptors, the emotional coding applied the following rules: with respect to the evaluation of the valence, the semantic part of the comments was used, whereas with respect to the emotional intensity, the valence from the semantic part of the comments was used and for the evaluation of emotional intensity, the code was focussed on the formal part of the comment. In particular, we start from the "typographical energy" cues used both in CMC (Sproull and Kiesler, 1986) and in the psycho-linguistic literature (Caffi and Janney, 1994; Poggi, 2001; Bonelli, 2015). The "hostile" cues consist of the following: the length of the comment (more than 50 words calculated on the basis of the mean length of the comments); typographic extremes marked by exclamative or interrogative points; words with capital letters; use of negative similes; bold or coloured letters (e.g. red); and use of insults, profanities, swearing and derogatory names.

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Criteria</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>No hostility</td>
<td>A short length (less than 50 words) without any cues or explicit expressions of hostile emotions</td>
<td>I am sorry about this situation, but I do not think the solution is to come here [...]</td>
</tr>
<tr>
<td>Annoyance</td>
<td>At least one of the described cues and expression of emotion in words like (annoyance, bother, trouble, discomfort; problem)</td>
<td>No, no, we must not forget our migrants, but today unfortunately are different times since then. Europe must do its part all!</td>
</tr>
<tr>
<td>Irritation</td>
<td>Two or more cues and clear semantic expression of anger</td>
<td>Gianni, don’t make me angry, not fall in the obvious bourgeois respectability [...] then give back to the lands to the Indians who were taken away by force [...] We were not welcomed in luxury facilities, heated and fed with other people’s money [...] please do not make FUTILE COMPARISONS!</td>
</tr>
<tr>
<td>Contempt</td>
<td>Presence of cues 6 or explicit expression of hostility and hate through different speech acts plus typographical markers</td>
<td>We have to SHOOT OFF!! is all delinquency extra we have to maintain!!!! Fuckkkkkkkk!</td>
</tr>
</tbody>
</table>

### Table II. Codebook of hostile emotions

<table>
<thead>
<tr>
<th>Emotions</th>
<th>Criteria</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>No hostility</td>
<td>A short length (less than 50 words) without any cues or explicit expressions of hostile emotions</td>
<td>I am sorry about this situation, but I do not think the solution is to come here [...]</td>
</tr>
<tr>
<td>Annoyance</td>
<td>At least one of the described cues and expression of emotion in words like (annoyance, bother, trouble, discomfort; problem)</td>
<td>No, no, we must not forget our migrants, but today unfortunately are different times since then. Europe must do its part all!</td>
</tr>
<tr>
<td>Irritation</td>
<td>Two or more cues and clear semantic expression of anger</td>
<td>Gianni, don’t make me angry, not fall in the obvious bourgeois respectability [...] then give back to the lands to the Indians who were taken away by force [...] We were not welcomed in luxury facilities, heated and fed with other people’s money [...] please do not make FUTILE COMPARISONS!</td>
</tr>
<tr>
<td>Contempt</td>
<td>Presence of cues 6 or explicit expression of hostility and hate through different speech acts plus typographical markers</td>
<td>We have to SHOOT OFF!! is all delinquency extra we have to maintain!!!! Fuckkkkkkkk!</td>
</tr>
</tbody>
</table>

5.3 Results

Unethical orientation and hostile emotions. A total of 10,137 comments were codified. This revealed a polarised discussion, including, on the positive side, 5,047 prosocial comments (49.78 per cent) and, on the negative side, 4,551 comments based on MD (44.89 per cent) and 167 comments featuring discrimination (1.64 per cent). In addition, 372 comments were coded as "personal venting", personal stories in which the users complained about how difficult it is to cope with life, without taking an explicit position towards hosting immigrants (3.67 per cent). As to hostile emotions, irritation was the most frequently expressed (40.4 per cent), followed by no expressed hostility (24.1 per cent) and annoyance (22.5 per cent). The levels of irritation were higher across the analysed period (from 21 to 27 April 2015) compared to other types of hostile emotions ($\chi^2_{(18)} = \text{267,697}, p < 0.001$).

With regard to gender differences, the findings were in agreement with the classical literature on emotions suggesting that women tended to express fewer hostile emotions than men (Eagly and Steffen, 1984; Deaux and Major, 1987; Chaplin et al, 2008). In fact, annoyance (53 per cent vs 47 per cent), irritation (53 per cent vs 47 per cent) and contempt (63 per cent vs 37 per cent) were all significantly higher in men users ($\chi^2_{(6)} = \text{50.72}, p < 0.001$).

The relation between type of comment and hostile emotions is represented in Table III, where it emerges that prosocial comments and personal vents were expressed without hostile emotions (89.5 and 30.7 per cent, respectively). This contrasts with MD comments and those featuring explicit discrimination, where irritation was higher (40.7 and 54.5 per cent, respectively) ($\chi^2_{(9)} = \text{4580.33}, p < 0.001$).
Online MD. The analysis of frequencies (Table IV) indicated displacement of responsibility as the MD mechanism that appeared most frequently (29 per cent). It also received the highest number of “likes” (33.15 per cent). In this case, as reported in the examples below, Facebook users attributed their lack of agency in accepting/hosting immigrants to contextual factors that highlighted a sense of distrust towards authority (i.e. administrative authority, state, politicians and government), with recurrent themes such as, bad management of public affairs, lack of controls and procedures, no guarantees, the state has no ability to enforce laws, the state voluntarily creates these conditions for a war between the poor. This mechanism was used more by women than by men ($\chi^2 = 54.29$, $p < 0.001$), who instead tended to use responsibility diffusion. While women attributed the issue to more general negative conditions, men attributed it to more precise categories (i.e. rich people):

We must not accept it if there are no guarantees from the authorities, adequate controls.

It’s right not to welcome them when there are no suitable conditions for it.

Our country can’t afford it.

We have no serious politicians that understand what to do, but only people, among others who are incompetent, who are incapable of considering the dramatic situation of these poor human beings, and make useful and decent decisions for them and who come to us and tell us that we have to receive them.

The second most used MD mechanism was that based on the locus of “outcome”, namely the distortion of consequences (18.52 per cent, liked by 21.53 per cent). In this case, this mechanism was focussed on the worsening of social conditions if would-be immigrants were to be accepted. The following recurrent themes appeared in comments:

They steal our jobs, interfere with our comfort, only spread diseases, get what Italians are denied, are treated better than us; they destroy everything, impose their customs and traditions, come to sell drugs in our poor country.

<table>
<thead>
<tr>
<th>No hostility (%)</th>
<th>Annoyance (%)</th>
<th>Irritation (%)</th>
<th>Contempt (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrimination</td>
<td>9.58</td>
<td>8.38</td>
<td>54.49</td>
</tr>
<tr>
<td>Moral disengagement</td>
<td>24.17</td>
<td>22.56</td>
<td>40.75</td>
</tr>
<tr>
<td>Prosocial</td>
<td>88.54</td>
<td>4.68</td>
<td>3.33</td>
</tr>
<tr>
<td>Personal vent</td>
<td>30.73</td>
<td>29.38</td>
<td>29.92</td>
</tr>
<tr>
<td></td>
<td>56.71</td>
<td>13.67</td>
<td>21.95</td>
</tr>
</tbody>
</table>

Table III. Type comment × hostile emotions

<table>
<thead>
<tr>
<th></th>
<th>%</th>
<th>n</th>
<th>% likes</th>
<th>n likes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responsibility displacement (A)</td>
<td>29.00</td>
<td>1,220</td>
<td>33.15</td>
<td>5,036</td>
</tr>
<tr>
<td>Consequence distortion (O)</td>
<td>18.52</td>
<td>843</td>
<td>21.53</td>
<td>3,271</td>
</tr>
<tr>
<td>Blame attribution (V)</td>
<td>18.45</td>
<td>840</td>
<td>21.03</td>
<td>3,194</td>
</tr>
<tr>
<td>Dehumanisation (V)</td>
<td>14.61</td>
<td>665</td>
<td>8.54</td>
<td>1,297</td>
</tr>
<tr>
<td>Responsibility diffusion (A)</td>
<td>10.92</td>
<td>497</td>
<td>10.66</td>
<td>1,619</td>
</tr>
<tr>
<td>Moral justification (B)</td>
<td>7.47</td>
<td>340</td>
<td>3.25</td>
<td>494</td>
</tr>
<tr>
<td>Advantageous comparison (B)</td>
<td>0.55</td>
<td>25</td>
<td>1.67</td>
<td>253</td>
</tr>
<tr>
<td>Labelling (B)</td>
<td>0.48</td>
<td>22</td>
<td>0.17</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>100.00</td>
<td>4,532</td>
<td>100.00</td>
<td>15,191</td>
</tr>
</tbody>
</table>

Notes: A, agency; O, outcome; B, behaviour; V, victim

Moral disengagement and hostile emotions
They come here and find paradise, cigarettes in hand, and do nothing, complain about the food offered to them, they come here to lay down the law, sunbathing.

Taken together, the mechanisms focussed on the victim made up the greatest number of MD comments (33.06, 18.45 and 14.61 per cent for blame, attribution and dehumanisation, respectively). These were the harshest and most explicit mechanisms. Users who cast judgement on the victims assigned responsibility for their condition to them (including to their “clandestinity”: to be clandestine became a fault). Their comments suggested that immigrants were deserving of harm and to be denied help, until they were reduced to a status of sub-humanity. With blame attribution, users typically posted comments such as, “If we do not host them, it is their fault because they just claim things for their interests, they do not work and they do not respect our rules and norms”. With dehumanisation, recurrent themes were those that involved descriptions of immigrants as dependent people who abused Italians: parasites, luxury parasites, jailbirds, garbage.

Diffusion of responsibility, which appeared in 10.92 per cent of users’ comments, involved explaining unwillingness to host immigrants on the grounds that rich people, such as politicians, and in particular Gianni Morandi himself, did not help them. Thus, the responsibility for the immigrants’ condition was diffused to wealthy individuals, as in the following cases:

Rich people do not do it, why should we do it? Politicians can do it!

It is not permissible to accept them if they don’t, even those who could afford it financially.

It is reasonable for a country not to accept them if other European countries also do not accept them.

Finally, the less used mechanisms were those based on users’ behaviour (potential lack of assistance), which, in total, represented only 8 per cent of the whole corpus. These mechanisms were based on the “common good” and seemed to be more self-regulated, given that they appeared to be related to a user’s concern for his/her own reputation. With moral justification, omission was described as motivated by social and moral purposes:

It is right not to accept them, so as not to create false hopes and illusions.

[…] accepting them means just giving them the chance to commit crimes because there is no work.

With euphemistic labelling, users described their potential omission with euphemistic language in order to make it more morally acceptable. Hosting immigrants was linked to human business, cash machine and clandestinity:

Hosting is just a form of “do-goodism”, foolish and criminal.

With advantageous comparison, users compared their omission with worse and more flagrant conduct, as in the following case:

Better not to accept them (prevent them from leaving) rather than have them die at sea/have their deaths on our conscience.

**The relationship between MD and hostile emotions.** In order to test the discriminant validity, we ran a \( \chi^2 \) analysis (\( \chi^2 \) (121) = 520.27; \( p < 0.001 \)) by also adding the Cramér’s coefficient (\( \nu = 0.29; \ p < 0.05 \)), which, being low, together with a low contingency coefficient (\( r = 0.32; \ p < 0.05 \)), reveals a high level of independence between MD and hostile emotions.

Correspondence analysis is a particular type of principal component analysis. Its aim is mainly to provide a description of the underlying features that define each factor without presuming to confirm or deny a causal relationship between the chosen variables
This type of analysis was chosen to present the results pertaining to MD mechanisms and hostile emotions visually.

Correspondence analysis (Figure 1) on this data set extracted two main factors, which explain, respectively, the 77.55 and 19.62 per cent of inertia (variance). The first may be defined as hostility bias (low vs high), while the second is attentional focus (individual vs social/contextual). The first factor contributed substantially to the analysis (77.55 per cent) and, as reported in Table V, it is defined by an opposition between the first and fourth factors.

Notes: F1, factor 1: “Hostility bias” (low vs high); F2, factor 2: “Attentional focus” (individual vs contextual). F1 and F2 percentages represent the explained inertia (i.e. the variance of each factor). The ten red points represent 8 MD mechanisms plus personal vent and discrimination; the four green squares indicate the four type of hostile emotions
quarters of the graph (I and IV) and the second quarters (II and IV), which can be grouped as social media users who felt a higher level of hostile emotions. Such hostile emotions were reflected in comments about how dangerous immigrants can be. These comments exhibited explicit discrimination, dehumanisation, blame attribution (victims) and consequence distortions (outcome). Differently, comments which manifested lower levels of expressed hostility towards immigrants, tended to show more indirect (and more self-regulated) MD mechanisms such as displacement of responsibility (outcome), advantageous comparison and euphemistic labelling and moral justification (behaviour – in this case, their omission or lack of assistance). The diffusion of responsibility falls within the high hostility bias because this includes the people who tried to “solve” the problem by passing the responsibility to “rich people” and, in particular, to Gianni Morandi as representative of the wealthy class. Overall, in online contexts, emotional arousal, when it becomes progressively more intense, can be associated with cognitive distortions of negative effects that hosting can provoke and with the negative representations of the victims, as in the case of dehumanisation.

The second factor (inertia, 19.62 per cent) is the “attentional focus”, which involves users opposed to immigration who were personally or contextually/socially focussed; therefore, in the correspondence plan, quarter I–II (Figure 1) included users focussed on individuals, who, when they felt contempt, attributed responsibility to individual features of immigrants (“They come to steal”, “They are parasites”), or who, when they did not feel hostile emotions, considered it acceptable not to provide help on the grounds of their personal moral principles (“It is right to consider their future here”). On the other hand (quarters III–IV), social media users, when they expressed irritation, were contextually focussed and anticipated the potential social consequences (“They come here to change our religion”, “They steal our jobs”, “They spread diseases”). When they felt annoyed, they attributed external factors to their lack of help (“There are control procedures”, “Italy is a country with an economic crisis”, “There’s no work here”).

6. Discussion
This study, by adopting a psycho-lexicographical approach, involved the examination of the relationship between hostile emotions and MD mechanisms expressed during online interactions when people were discussing the hosting of immigrants. For this purpose, the case of Gianni Morandi was chosen, in which Morandi drew attention to the issue of hosting immigrants by trying to foster the similarity/closeness between the difficulties of past Italian immigrants and actual would-be immigrants from Africa. Morandi’s Facebook post, attempting to promote closeness between potential helpers (Italians) and those in need of help (immigrants) through social media, triggered a broad discussion (more than 11,000 comments in a week), characterised by hostile tones and different justifications offered for not accepting immigrants. The case was studied by integrating MD (Bandura, 1991, 2015) and hostile emotions theories (Izard, 1975; Poggi, 2001, 2008; D’Errico and Poggi, 2016). This study on hosting immigrants, which integrated these two theoretical lenses, yielded a more analytical understanding of the possible associations between affective states and self-exonerating mechanisms, by highlighting that, the more hostile emotions people express, the more they use MD mechanisms that deny the status of dignity to people in need.

6.1 Online MD
In contrast with previous studies (Faulkner and Bliuc, 2016; Connelly et al., 2016), which mainly focussed on incidents of real racists’ comments, a social media case was found to be suitable for interpreting the MD mechanisms in terms of an unwillingness to help, which is a potential omission towards immigrants. In line with studies on prosocial behaviour (Paciello et al., 2013) and ethical choices whose harmful consequences are not immediately
tangible over the long-term (such as ecological behaviours) (Bandura, 2007), the results suggest that MD mechanisms intervene in disabling the proactive route of moral self-regulatory processes, explaining why people do not behave ethically (omission), even when they may know it is right to do so. The results further highlighted how people, when they are oriented towards omission, use a broader range of MD mechanisms, at least much more than in the case of active harmful behaviour (e.g. racism; Faulkner and Bluc, 2016; Connelly et al., 2016). In particular, the ways in which people may use MD mechanisms in this case have their own peculiarities. Indeed, MD mechanisms may take different forms while maintaining the same function, that is, legitimising unethical orientations. For example, euphemistic labelling allows a cognitive restructuring of behaviour, since it makes a positive behaviour into something negative or socially and morally wrong (e.g. hosting is only a “cash machine for mafias”). Furthermore, distortion may not decrease the seriousness of negative behaviour but disparages or discredits the value of positive action or even transforms it into something negative (e.g. immigrant hosting is something harmful to citizens). Similarly, regarding agency, people might attribute to outside sources their inability to behave in an ethical manner (e.g. hosting is not the responsibility of individual citizens, but the state/authority is acting irresponsibly). Finally, with regard to the mechanisms that focus on the victim, people can consider others who need help as worthy or not worthy of being accepted within their social group (e.g. “They are parasites, do not deserve to be helped”). These mechanisms (attribution of blame and dehumanisation) represent those parts of the total disengaged comments (33 per cent) characterised by very aggressive and hostile contents.

Interestingly, the most widely used mechanism that received the most votes (“likes”) was that of displacement of responsibility (29 per cent of all MD mechanisms, with 33.15 per cent “likes”). This suggests a deep sense of mistrust towards institutions (as in the case of national and European authorities who do not guarantee respect for the “hosting procedures”, or “safeguarding the rights of citizens and immigrants”), and, therefore, MD has to be read in the socio-political context. These results clearly show how the “moral” approach to the study of online discussions about immigrants offers a chance to widen the perspective from a simple attack on the victim (as in the case of flaming and racism) to a more complex study of the attribution of responsibilities for the potential lack of help. For example, people with an unsupportive stance (locus on behaviour) may believe that helping can be counterproductive or that they can disengage themselves by pointing to others’ lack of help, with references to “corrupt government” and “unsupportive rich people” (locus on agency). These self-exonerating mechanisms can be strongly related to the potential helper’s affective state, as suggested by several studies (Paciello et al., 2013). For example, personal distress can favour it (Batson and Oleson, 1991), but the relationship remains “complex and multifaceted” (Eisenberg and Miller, 1987).

Hostile emotions and online MD. The study of affective and moral cognition is gaining more attention among researchers (Greene et al., 2001; Haidt, 2001; Manfrinati et al., 2013). However, in respect of MD, what has not been examined in depth is the relationship between different types of emotion and the activation of these self-serving mechanisms that allow individuals to bypass their moral control and behave in a more utilitarian and ego-centred way.

The relationship between MD and negative emotions has been suggested also by studies that analysed the mediational role of MD on types of aggressive and deviant conduct (Caprara et al., 2014; Fida et al., 2015; Rubio-Garay et al., 2016). In contrast with these previous findings, which have used mono-dimensional scales, this study highlighted that increasing levels of hostile emotions are associated with specific MD mechanisms. Thus, studying each MD mechanism separately demonstrated that, with greater intensity/activation of hostile emotions, the disengagement locus moves from the...
individual’s behaviour (in this case, that of the potential helper) to the target recipient of such behaviour (the person in need).

In particular, it was observed that, in the relationship between hostile emotions and MD mechanisms, two factors can identify the different connotations of arguments representing an unethical orientation towards immigrants: the “hostile bias” (low to high; Factor 1) and the “attentional focus” (individual vs context; Factor 2). In relation to a low “hostile bias” with an “individual” focus, the MD mechanisms were almost absent, with the exception of moral justification, which here deflected attention away from the individual towards misleading personal values (“Do not deceive other people”). Otherwise, when the hostile tone increased with annoyance, the MD mechanisms that operated at the behavioural level and on an agency level were strongly supported by context-specific arguments. In particular, omitting to offer help was mostly seen as a result of external conditions: as in the case of responsibility displacement, personal responsibility was deflected to absent institutions, practical and social conditions were deemed not conducive to good behaviour or, with respect to labelling, hosting became unethical behaviour because it was only another form of encouraging deviant support systems.

Evaluation of hosting became even more hostile (in the case of irritation) and contextually focussed in relation to the distortion of consequences. Distortion led people to see the hosting outcomes as harmful for their own community and overly beneficial for immigrants. There were many references to contextual factors (e.g. hotels, comfort and tax advantages) characterised by an intense sense of social injustice that resulted from a paradoxical representation of immigrants as “lucky”. An irritated tone characterised the mechanism of diffusion of responsibility, directed towards a specific social category of “wealthy people”, with Gianni Morandi as their representative, who, despite their financial ability to do something, remained inert. The tones associated with this mechanism were activated to an unexpectedly great extent, possibly because people represented in their mind not only immigrants but also their defender (Morandi), making him worthy of social envy.

The unethical orientation became unfair in relation to contempt. In fact, the disengagement mechanisms used were dehumanisation and the attribution of blame, together with explicit discrimination (Kelman, 1976; Bar-Tal, 2000). Here, contextual references disappeared from the comments, which were generally poor in terms of content. The focus of attention was on more individual aspects of the victims (e.g. negative disposition), and in no way was consideration given to their social, cultural and human conditions. The MD mechanisms were, thus, based on a few impressions of immigrants’ personalities as stable individual traits, stereotyped and dehumanised (Volpato, 2014) with aggressive language (D’Errico et al., 2014).

Consequently, these results showed a broader scope of the possible disengagement mechanisms in discussions concerning the hosting of immigrants, not only those that were strongly directed towards the victims (racism), but also the more indirect and subtle ones. Furthermore, the analysis of co-occurrences with hostile emotions could explain not only how the expressed contempt can contribute to dehumanisation (Kelman, 1976; Bar-Tal, 2000) but also how a person who expresses his/her comment without hostility or at most an annoyed emotional tone can also distort his or her own responsibilities. More generally, the results of the present study contribute to the general literature on MD mechanisms by focussing on the role played by hostile emotions and highlighting how one’s emotional aspect can create a “moral” distance from the needy other via a distortion of one’s moral image (i.e. “He is not worthy of being helped”). Furthermore, the present contribution, which underlines the added value of considering distinctly MD mechanisms (usually studied as an aggregated measure) aids understanding of the possible and tailored strategies people use to re-orient the “moral compass” (Moore and Gino, 2013). For example, in dealing with
people who displace their own responsibility, it can be useful to promote both personal and collective accountability by pointing to virtuous examples; in dealing with people who tend to dehumanise others, it can be useful to show immigrants by giving them a face and an identity and thus promote empathy. In addition, the results suggest that emotion regulation can be another impulse that can intervene to hinder unethical orientations, especially dehumanisation. Overall, the results show that moral dysregulation can be associated with difficulties in emotional regulation.

7. Conclusions

Social networks can trigger and spread hostile emotions and MD mechanisms that feed the spiral of incivility. This contribution is meant to represent the first investigation and reflection of the role that these dimensions may jointly have with respect to the spread of increasingly harmful and unethical orientations. Understanding them interactively can help to prevent and manage such unethical orientations. The results suggest that in an environment in which the representation of reality can be misrepresented or distorted and, based on exclusively negative beliefs and partial understanding of a complex phenomenon, such as immigration, moral regulation may be seriously undermined. It is, therefore, important even in these contexts to identify “external triggers that provide a crucial route for activation of self-regulatory mechanisms” (Runions and Bak, 2015, p. 402).

The present study was aimed at analytically understanding the synchronic relation between expressed hostility and online MD mechanisms implied in online discussions, without focussing on the causal relation between them. It is likely that expressed emotions and MD mechanisms are mutually influenced in mediated computer environments where people can easily reinforce their beliefs (e.g. confirmation bias; Quattrociocchi et al., 2014) and be contaminated by hostile emotions (Kramer et al., 2014). Furthermore, these dimensions can be intercepted in social media by investigating the triggering factors. The study of the comments suggests that further investigation is required into the contextual factors on which many behavioural and consequence distortion mechanisms rest and into the apparent decrease in the sense of personal agency. For future study, the investigation of the following factors would be useful: the role of “ethical” communication models (i.e. by studying “ethical frames” starting from the “classical” media) and the transmission and acquisition of beliefs through social media processes that could create the conditions for the legitimisation of behaviour usually considered unethical; the role of individual differences in the accessibility of hostile cognitions and emotions, to the detriment of the acceptance and integration of needy minority groups. One of the major limitations of the present study is the absence of access to social and individual factors that may contribute to moderating this process of disseminating and strengthening unethical orientation. Future studies could help in this regard by using different methodological approaches (i.e. experimental and physiological studies) and tools of detection built specifically for the study of psycho-social phenomena in social media.

Moreover, in validating and enlarging the referenced theoretical models, it is possible, as suggested by an anonymous reviewer, that another limitation could be the adherence to theory, which may have prevented new insights from emerging based on the analysis of the data set. However, the adoption of a socio-cognitive framework was deemed necessary to understand the relevance and opportunity of studying MD together with hostile emotions in an online setting. Indeed, as also argued by Bandura (2015), “Moral battles are now being waged transnationally by Internet-savvy participants sharing morally relevant information via blogging, podcasting and texting with unfettered, wireless devices. These globalised modes of influence are being used both in the service of MD and to expose the use of such practices” (pp. 7-8). Understanding how MD mechanisms are related to hostile emotions could provide scientific knowledge useful for preventing and counteracting the suspension of morality in an online setting.
At application level, this study sheds light on possible areas where education can intervene by identifying personal or contextual representations of unethical orientation. Also, in technological contexts, the contribution can be useful for implementing algorithms to identify users’ online self-regulation that can be detected through the hostile lexicon and morally disengaged states (Scheuermann and Taylor, 1997). For example, in the case of “cognitive distortion”, the system can suggest “social advertising” banners restoring the value dimension or giving the viewer’s alternative information on immigrants. This can be a possible social media strategy in the light of recent research on “moral contagion” (Brady et al., 2017), which shows how emotions and moral dimension interact.

Note
1. People can freely comment on Morandi’s Facebook page (as a public character, he leaves comments completely open to both followers and non-followers; moreover, he has a very large number of followers – more than 2,390,000).

References
Bandura, A. (2006), Manual for Coding Modes of Moral Disengagement, Stanford University, Stanford, CT.


Further reading


About the authors

Francesca D’Errico is Assistant Professor of Social Psychology at Roma Tre University, Fil.Co.Spe Department (Philosophy, Communication and Visual Art). Her main research interests concern aggressive verbal and non-verbal behaviours like discrediting and insulting within political and social media communication. She also works on emotional expressions of conflictual contexts by exploring cognitive and communicative ingredients as in the case of bitterness, pride and acidity as indirect way of expressing one’s own sense of injustice. She authored/co-authored more than 80 studies published on journals and chapters, and she also co-editor of two Springer Books on Multimodal approach to the study of conflictual political communication. Francesca D’Errico is the corresponding author and can be contacted at: fderrico@uniroma3.it

Marinella Paciello is Associate Professor of General Psychology at International Telematic University UNINETTUNO, Faculty of Psychology, Italy. Her recent research interests are mainly related to cross-sectional and longitudinal impact of personality dimensions on prosocial and antisocial behaviours, especially in clinical and organisational settings. In particular, within a social cognitive perspective, she has investigated the role of moral disengagement in fostering unethical conduct and of self-efficacy beliefs in promoting positive adjustment. She authored/co-authored several studies in peer-reviewed journals, resulting from different interdisciplinary collaborations aimed to understand ethical and moral functioning.

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
The bidirectional mistrust
Callers’ online discussions about their experiences of using the national telephone advice service

Annica Björkman
Department of Public Health and Caring Sciences, Health Services Research, Uppsala University, Uppsala, Sweden and
Martin Salzmann-Erikson
Faculty of Health and Occupational Studies, University of Gävle, Gävle, Sweden

Abstract
Purpose – The purpose of this paper is to explore and describe online communication about the experiences and attitudes toward Swedish Healthcare Direct, a national telephone advice nursing (TAN) service.
Design/methodology/approach – A descriptive research design was adopted using a six-step netnographic method. Three Swedish forums were purposefully selected and data from the virtual discussions were collected.
Findings – Three themes emerged: expectancy and performativity of the nurses, absurdity in accessibility and the scrutinizing game. The most prominent finding was the scrutinizing game, which included aspects of bidirectional mistrust from both nurses and callers. Another salient finding was the attitudes that callers held toward nurses who used a technique interpreted as “passing the buck.”
Research limitations/implications – The use of a netnographic method is novel in this area of research. Consequently, the body of knowledge has regarding telephone advise nursing service has significantly been broadened. A limitation in this study is that demographic data for the posters are not available.
Practical implications – Bidirectional distrust is an important issue that must be acknowledged by TAN services, since it might damage the service on a fundamental level. Healthcare providers, politicians, and researchers should account for the power and availability of virtual discussions when seeking consumers’ opinions and evaluating the quality of the care provided.
Originality/value – This analysis of the ongoing discussions that take place on the internet provides insight into callers’ perceptions of a national TAN service. The bidirectional mistrust found from both the nurses and the callers might be a threat to callers’ compliance with the advice given and their care-seeking behavior.

Keywords Consumer attitudes, Nursing, Distrust, Internet forums, Antagonistic effects, Cyber nursing

Paper type Research paper

Introduction
Telephone advice nursing (TAN) is one way of steering patient flows through the limited resources of health care organizations. Each caller is triaged by a registered nurse and, if appropriate, referred to the correct level of care. TAN is common within several countries such as the UK, Canada, the USA and the Netherlands (Knowles et al., 2014; Snooks et al., 2008). NHS (2018) 111 in the UK answered 1,564,236 calls in March, with an average of 50.5 thousand calls per day (NHS 111). However, the way it is performed and organized varies depending on the country’s healthcare organizations (Hansen and Hunskaar, 2011; Knowles et al., 2014; Rutenberg and Greenberg, 2012). The Swedish TAN service, Swedish Healthcare Direct (SHD), was launched in 2003 and was inspired by NHS 111 in the UK. SHD is one of Sweden’s largest healthcare providers, receiving 5.5 million calls (i.e. patient contacts) each year. It is reachable 24 hours a day, every day of the year, and access is free of charge. TAN in Sweden is regulated in the same way as other forms of Swedish healthcare. SHD’s mission is to produce “good health and healthcare on equal terms for the entire population” and to work to “prevent ill health” (HSL, 1982/1999, p. 763; Patient Act, 2014, p. 821).

Online communication has shown to tremendously impact people’s attitudes and apprehensions, from the influence of self-brand image (Islam et al., 2018) to major governmental changes (Howard and Hussain, 2013). Existing research within the field of
TAN has focused on the aspects of callers’ expectations, experiences (Kaminsky et al., 2013), and satisfaction (Rahmqvist et al., 2011), although with some methodological limitations. Furthermore, previous research on the subject has been mostly restricted to self-reports and qualitative interviewing methods (Kaminsky et al., 2013; Hansen and Hunskaar, 2011). The results of these studies have shown callers to hold a predominantly positive view. However, much uncertainty still exists regarding how callers perceive the quality of care, their experiences of TAN, and their attitudes to TAN. In contemporary society, the internet is increasingly being used to discuss individuals’ experiences and attitudes toward numerous fields of interest, including healthcare. The information which is shared in internet forums about TAN can be a powerful source of insights for professional, theory, and organizational development. The intention of this paper is to extend this knowledge through a previously unused methodology in the field of TAN.

Literature review

Callers’ perceptions of TAN

Callers’ perceptions of the service have previously been studied using self-reported data covering callers’ expectations, experiences, and satisfaction. Many callers were positive about the idea of receiving healthcare advice via telephone and reported that they appreciated receiving adequate advice from RNs in addition to reassurance, support and kind respectful treatment (Kaminsky et al., 2013; Strom et al., 2009; Winney et al., 2014). A study using a sample of 517 callers showed that three factors – interaction, service, and outcome of call – influenced overall caller satisfaction, and that overall satisfaction with the service was high (Rahmqvist et al., 2011). It is common for parents to call the service on behalf of their children, and Kaminsky et al. (2013) report how parents appreciated when telenurses invited them to call back if in doubt. Many callers already know they want a referral to a certain level of care before calling SHD, and hence have a “hidden agenda” for their call; a perception of not being well-treated has been shown to increase callers’ desire for healthcare appointments (Rahmqvist et al., 2011; Winney et al., 2014). A comparison of three groups of callers found that callers who received a lower recommendation of care than expected gave significantly lower ratings on three questions concerning the nurse–caller encounter, compared to controls (Rahmqvist et al., 2011). Callers’ compliance with advice is an interesting issue, since the TAN service has limited value to healthcare economics if the callers do not follow the advice they are given. An interview study showed how callers’ compliance with telenurses’ recommendations, as well as their satisfaction with the service, was affected by the nurses’ communication skills (Kaminsky et al., 2013). Parents also described how they relied to a large extent on the RN’s advice, and most parents stated they had learned something from their calls (Kaminsky et al., 2013). Furthermore, callers stated that they were experts on themselves (Winney et al., 2014) or their own children (Kaminsky et al., 2013). These expressions of self-determined expertise are rapidly growing among healthcare seekers and constitute a pedagogical challenge for RNs working in TAN services. Callers’ expectations for the service to provide “instant access” have been shown to contribute to feelings of stress among RNs working in TAN services. Nurses said that when taking this kind of call, they might simply accept a caller’s idea of what was wrong, rather than gathering sufficient information for a proper assessment. Demanding, angry callers could also be perceived as a threat to safety. Some conditions were described by the nurses as almost impossible to assess over the telephone (Roing et al., 2013). In contemporary society, patients are increasingly well informed about their rights within the healthcare system and are therefore able to place more active demands on healthcare personnel and on the system (Longtin et al., 2010). A possible reason for this is the easy and instant access to information on the internet regarding healthcare and disease management (Medlock et al., 2015; Moreland et al., 2015).
Another aspect of online counseling services is their accessibility. Cipolletta and Mocellin (2016) state that online services such as SHD might be considered useful in the public field, as they increase accessibility and provide benefits for the community by taking care of the less-demanding parts of healthcare, such as providing the first means of access to traditional counseling and care. Patients’ use of the internet can affect the nurse-patient encounter in several ways. When patients collect healthcare information online about alternative treatments or experimental therapies, they simultaneously reconfigure the lay–professional divide in medicine (Stacey et al., 2009). Patients become “armed” with information and empowered to take ownership of their health and of communication with the healthcare providers (Fox et al., 2005; Korp, 2006). As patients increasingly consider themselves consumers rather than patients, they are more inclined to refuse to accept the traditional paternalism, even though the healthcare provider still has the power (Lupton, 1997). Patients are more likely to “shop around”: that is, to actively evaluate the services offered, and to go elsewhere if they find the “commodity” offered “unsatisfactory.” In what ways “patients as consumers of care” communicate online about their experiences of TAN has not been elaborated in previous research. However, it is likely to argue that their communication may impact others’ view of TAN, just as online communication can impact other governmental institutions.

The power of online discussion forums

Discussions on online forums have been shown to help users express themselves more freely than in face-to-face contexts (Cipolletta and Mocellin, 2016). Similarly, previous research has acknowledged the strong impact of consumers’ ability to voice their attitudes toward a brand or company in the form of user-generated electronic word-of-mouth (Brown et al., 2007; Chen et al., 2017; Xun and Guo, 2017). Patients’ roles as consumers are reinforced in several ways, including by HSL 1982/1999, which states that patients should be encouraged to make decisions regarding their care. Brown et al. (2007) have shown how online consumers are active and judicious and hence possess valuable information regarding both the design and usability of a product. Online communities have been shown to have the potential to affect the attitudes and experiences of consumers. Within online social networks, information such as “testimonials” from other participants gains credibility within the community and becomes “the truth.” According to Hotwire (2006), blogs are regarded to be the second most trusted information source. As stated by Ketelaar et al. (2015), negative opinions have more impact than positive opinions in online discussions. Despite the fact that TAN plays such a central role in SHD, a search of the literature revealed that relatively few studies have evaluated the service from the viewpoint of the callers.

Research question

RQ1. What are the experiences and attitudes toward the Swedish national TAN and how is it discussed in online communication?

Method

A descriptive research design (Patton, 1990) was chosen in order to gain an in-depth understanding of a specific social phenomenon. A review of the existing literature revealed a lack of studies investigating communication among people who visit an internet-based forum to discuss their experiences of and attitudes to using the TAN service. In studying social phenomena, it is not possible to capture an absolute truth, although the phenomenon
of study can be captured and described in one of many plausible ways. Thus, the
descriptions must take into account the fact that the callers’ experiences and attitudes are
influenced and colored by their social, historical, and cultural context (see for example
Glaser and Strauss, 2017; Miles, 1994). In the past, cultures and interest groups were located
geographically in the same area. However, with the development of computer-mediated
communication, cultures are no longer bound to a spatial spot, and thus cyber cultures have
developed (Boyd and Ellison, 2008; Kozinets, 1998). The present study’s epistemological
approach is in line with Halford and Leonard (2003): “Not only do people make spaces, but
spaces may be used to make people” (p. 202). Thus, online communications not only
translocate sites of the social into a cyber-culture (Kozinets, 2015), but also form, construct,
and reconstruct the identities, attitudes, values and norms of posters (i.e. the authors of
electronic posts in a forum). A very concrete example of the extraordinary power that online
communication has to launch groundbreaking changes is the Arab Spring of 2011. People
used Twitter and other computer-mediated communication tools and platforms to tell
stories, thus spreading ideas of freedom and democratic values (Howard and Hussain, 2013).
There are numerous examples in the literature of research that adopts the basic principles of
qualitative and quantitative approaches to study online communication. Several different
terminologies are used for this, including internet studies (Wellman, 2004), digital
anthropology (Miller and Horst, 2013), cyber-ethnography (Lee, 2017), virtual ethnography
(Hine, 2000) and netnography (Kozinets, 2015). Kozinets criticized the blunt and analog use
of netnography in relation to Hines’ (2000) virtual ethnography, with the main critique being
that netnographic studies aspire to a depth of inquiry into the cyber cultures, whereas the
epistemology in Hines’ method is superficial. Previous scholars have highlighted the
potential usefulness of netnographic studies. For example, the tourism researchers Zhang
and Hitchcock (2017) analyzed blogs written by Chinese female tourists while traveling in
Macao and concluded that a gendered gaze is a socially contingent phenomenon. In addition,
market researchers have also used netnography to map the identity of people who engaged
in online discussion using a specific hashtag (Cuomo et al., 2016). In the present field of
research, nursing science, Salzmann-Erikson (2015) conducted a netnography called “Space
Invaders” to analyze how artifacts in nursing home environments exercise disciplining
structures over the elderly. The netnographic approach was considered suitable for this
study for two reasons. First, as stated by Langer and Beckman (2005), netnography is a
suitable method when it is difficult to recruit participants to a study. The availability of
web-based discussions provides immediate access to data without disturbing the
presumptive participants. Second, conducting an observational netnographic study makes it possible to access the “naturally occurring setting” (Brewer, 2000), and hence to
study posters’ behavior in an everyday context which was not influenced by the
researchers. Hence, it was also possible to engage in an analysis of the data that gave
meaning to the actions as ascribed by the posters (Hamarsley and Atkinson, 2007). The
main principles and steps of a netnographic study were used with the guidance of Kozinets
(2002). The suggested steps involve the following: definition of the research field, research
ethics, community identification and selection, community observation and data collection,
data analysis, presentation of the results, and interpretation and discussion of the results.

Data collection
The first part of the data collection consisted of locating the forums which included
discussion relevant to the study’s purpose, or in the terms of Kozinets (2002), definition of
the research field. We discussed research ethics and agreed that protection of posters was
important. The ethical discussion was also involved when forming the inclusion criteria of
the threads as: being considered relevant to the research question, including posts and
discussions that described a self-experienced situation of being in contact with SHD and
being available in a publicly accessible venue, meaning that the forum should be observable by anybody with no requirement for registration, login, or password. The exclusion criteria were: official healthcare web sites, statements by persons in their roles as professionals, and second-hand reports. In the next phase of community identification and selection we used the search terms “1177” and “forum” in the largest search engine, Google (www.google.com), which gave 710,000 hits. Numerous hits originated from the SHD website (www.1177.se), and so “site: www.1177.se” was added to the search string in order to exclude these; this reduced the number of hits to 698,000. Given the results of the search, both authors engaged in discussions about which forums should be included out of the many that were found. In order to obtain an adequate sample, it was agreed that three largest forums would be able to provide rich content. Due to the integrity aspects of the posters, we chose to keep the names of the forum cloaked. Regarding community observation and data collection, all three forums had a search facility, which was given the search term “1177” with the results sorted by “relevance.” The purposeful sample from the forums was copied and pasted into a separate Word document holding 230 pages (Forum A: 68 pages, Forum B: 70 pages, and Forum C: 92 pages) (Times New Roman font in 12 pt and set to A4 page size with 1.5-line spacing).

Data analysis
In concordance with Braun and Clarke (2006), the epistemological stance of this study was to conduct a latent content analysis; that is, to “examine the underlying ideas, assumptions, and conceptualizations” (p. 84). This decision made it possible to interpret the data on a level beyond the manifest level of the written words and sentences; hence we conducted a latent content analysis. First, the data were skimmed and read repeatedly in order to get an overall sense of the text. In the next phase, the whole data set was coded line-by-line, and segments were tagged and labeled (Miles, 1994). The material was scanned for data that would give an understanding of the online communication. Recurring words and phrases were noted down in a separate column next to the codes. Codes were grouped according to their similarities and differences, and then categories were constructed to provide an analytic description in accordance with the purpose of the study. Finally, the authors discussed the accuracy of the categories as the categories were compared with the raw data; thus we agreed to present the results in three categories. The presentation of the results was an abstracted text of our latent interpretation and strings of text were chosen that would provide good examples of each category.

Ethical considerations
All the data originated from open-access web pages, and none of the forums required any login or password. Moreover, no interaction with the forum posters took place, meaning that this cross-sectional observational study does not constitute human subject research. However, out of respect for the posters, their nicknames and the specific forum names have not been made public. Fictional nicknames are used below in order to make the excerpts in the results more vivid.

Findings
Callers who had contacted SHD used the service for a number of reasons; for example, a suspected tick bite, high fever, headache, or stomach pain. In some situations, they had called for help with their own problems, while on other occasions, they had called concerning problems related to their children’s or spouse’s health. Taken together, the findings of this study present an analytic description of the callers’ different experiences of seeking advice and communicating with SHD. Most posters seemed to be familiar with how to get in contact with SHD, as the service is the first contact with the healthcare system.
However, they had different attitudes and preconceptions toward the service. The callers seemingly wanted to receive credible and qualified support and professional advice from the nurses, but many of the forum posts bore witness to the opposite. Hence, even though the data revealed experiences of high-quality healthcare advice in the phone calls, many postings were about mistrust and complaints about indifferent nurses. A complex and bidirectional struggle was evident throughout the analysis. The results are presented in three themes due to their similarities and differences: expectancy and performativity of the nurses, absurdity in accessibility, and the scrutinizing game.

**Expectancy and performativity of the nurses**

In general, the posters were positive toward SHD as a public service organization, since it was accessible around the clock. They also encouraged others to call SHD when symptoms were discussed in the forum; one poster wrote:

> Otherwise, call the healthcare advice line if you’re worried. If you need help and advice you can always call the healthcare advice line, if you think they’re giving you the ‘wrong’ advice, tell them, and maybe you’ll get better help. Hope it all works out. (hummerniskara)

Where we are, the healthcare advice line is great, I’d rather call them than my primary care center. (josefinjanzon_81)

The posters said that they called SHD in situations when health advice was needed. One interpretation of this could be that initially there is a fundamental trust in SHD, since it is a public service healthcare organization that anyone can call and consult a nurse around the clock. However, such trusting attitudes were often expressed prior to the actual calls. Those who reported retrospectively about their contact with SHD expressed a more skeptical attitude, and this skepticism was often directed to the individual nurse who received their call. There were several postings which began by praising SHD and then continued with a subordinate clause beginning with “but.” One excerpt demonstrates this skepticism:

> The idea of the healthcare advice line is good, but in practice it’s a disaster. (mother_of_three)

However, other posters adopted a defensive role toward the critical posters, and demonstrated a humble attitude and forgiveness of the nurses’ shortcomings. The attitude of the individual nurse was highly important for the callers’ experiences:

> Since they obviously didn’t want to reply to my email, I just called them and got to talk to what I believe must be the world’s grumpiest old hag […] The only answer I got was: ‘Don’t you have a midwife there? Well, yes, I do. Well, go there then!’ [a/n: the poster recalls the conversation between herself and the nurse at SHD] The only answer I can give you is that you have to take a pregnancy test. How the hell can you be so grumpy?! I’m in so much pain today and the only thing she does is to be grumpy and basically call me stupid for calling […] In my ignorance, I thought the healthcare advice line was there to give you some idea about what to do, not dismiss you for reaching out to them! (kimii123)

Another poster had a similar experience:

> Met a real fucking grumpy cow at the other end of the phone who told me these were common pills for adolescents to abuse. And she gave me what I’d call a real telling off. (11anonymous)

There were several examples of calls where the posters had perceived nurses as unfriendly, arrogant, and disciplining. This could mean that when the nurses did not treat the callers in the respectful way the latter would expect from a public service and did not act in accordance with the callers’ expectations of how a nurse “should” address their problems, a complex and bidirectional struggle occurred.
Absurdity in accessibility

As demonstrated in the previous category, there was a fundamental trust in SHD, and one reason for this was that it was accessible around the clock. This accessibility was associated with having actual access to a consulting nurse. However, there were several postings describing disappointment in terms of accessibility. One poster wrote:

I've been trying to call all afternoon, and it's constantly busy. (lillan_87)

Others expressed their disappointment with the long telephone queues:

He was panicking and me and Jimmy were really frightened. 17 calls in the queue at SHD and a wheezing son. We went in, without waiting for SHD. (nineball_To_LUCK)

Even though most callers acknowledged that they had got in contact with a nurse when calling SHD, they criticized the accessibility of actual health advice due to receiving seemingly meaningless advice from the nurses. One poster even criticized the nature of the service:

What the hell are they even there for? A few times I've used this service and always been given the exact same answer, if the problem doesn't go away, see a doctor. This is something that everyone I've spoken to has experienced, too, they don't fulfill any actual function at all, and this is just an unnecessary job to reduce unemployment? (lannapatanna_1986)

My experience of the healthcare advice line is that it's a kind of 'first line of support' and its only purpose is stopping people from going to the ER that's already overcrowded. They've never given me any advice other than wait and see. Once I was bleeding pretty badly and needed stitches. (sore_face)

The callers questioned the very existence of the service. This can be interpreted as meaning that they sometimes felt SHD to be nothing more than a “false notion” in order to make citizens feel safe in the knowledge that they always have access to a consulting nurse. The false notion originates in the fact that, even though SHD is accessible per se, the service did not meet their expectations and did not provide access to advice. As seen in several excerpts, the callers perceived nurses as using delaying tactics or “passing the buck” to other healthcare services. One poster wrote that they wanted to be advised about their current symptoms, but instead they were given the advice to wait and see or to call their healthcare center the following day:

What's the point of calling them when it only ends up in 'Well, I don't know, but go to your healthcare center' [the poster quotes the nurse]? (mittnorland1)

The service was also criticized for it is extremely unbalanced “either/or” nature. Either the nurses neglected and downplayed the caller’s symptoms, or they overreacted and told the caller to go to the ER. One poster wrote about this phenomenon:

There's no point calling SHD. They send you to the ER where they yell at you for being stupid enough to listen to them. SHD is a big problem and seems to be at war with the ER. (galen_panna)

Another aspect of the absurdity in accessibility is what can be seen as "the safe approach." Callers had found that nurses listened to them but were reluctant to provide any analysis of the issues:

The person at the other end of the phone may well have the ability to make some guesses, but chooses to withhold this information because of policy. I completely understand that it might be difficult and they go for an idiot-proof approach [...] but I would have preferred being given the information at my own risk, so to speak. (anonym_1991)

When nurses adopt the safe approach, this might lead the callers to experience a sense of being diminished as an intellectual person and being left out of their own care, since they
“know” that the nurse has a whole repertoire of knowledge that is being kept from them. Hence, the caller’s descriptions of symptoms are manifested in the discussion, but the outcome of the nurse’s analysis remains the property of the nurse.

The scrutinizing game
The posts about callers’ experiences revealed complex communication patterns in the interaction. This study did not aim to perform an in-depth communication analysis, but throughout the data a power struggle became apparent, which must be acknowledged and elaborated on. There was bidirectional skepticism. The posters expressed doubts about the nurses’ advice, competence, and credibility. One wrote:

When you're advised to take two paracetamols and go to bed. Not go into the ER. When I was feeling really bad, and called them and described my symptoms, that's the exact advice I was given. The situation ended with my husband more or less forcing me into the car and driving me to the hospital. By then, my lips were purple and I was having trouble keeping my balance. Once there, they found that both my lungs were filled with 100 s of small blood clots. If I'd waited another hour, I wouldn't be alive today. I DON'T trust them. (snuffan)

The posters presupposed that the nurses were able to provide answers immediately, and criticized the nurses in situations when they seemed to be finding the answers on Google:

And seriously, are they real nurses who take the calls at SHD? I almost think it sounds like they're googling every question they get. (@tummen_upp)

In that situation, it’s quite boring to sit and listen to a nurse reading aloud from the same web page I just read myself. (CwC)

In this sense, the callers scrutinized and questioned the nurses’ education and competence. These postings often fueled other posters to write about similar experiences, but on rare occasions, some poster was of an opposing opinion:

As for myself, I've actually never come across a rude nurse, either at SHD or at the hospital […] you're probably just an awkward bastard (anonym_1313).

Just as the callers scrutinized the nurses, the callers wrote about their experiences of being scrutinized themselves by the nurses. Several posts bore witness to situations when callers experienced nurses who questioned their symptoms and credibility. These negative experiences of not being adequately assessed by the nurses at SHD led to the posters expressing dissatisfaction and distrust. One caller, who had been scrutinized about claiming a fever without being able to specify the degree, quoted the nurse:

SHD (in an annoyed voice): But then how can you know that you’re running a temperature? Me: For fuck’s sake, I've had a temperature before and I know what it feels like to have a temperature! (anonym_1313)

It’s real more the feeling of being treated like a child that annoys me. (Jokkmokk12)

One poster, who was a newcomer to the region, wrote about a call to SHD regarding a severe toothache. The poster had asked the nurse for advice about pain relief and dentists in the area, and received the following reaction:

Then she starts questioning me about why I'm in a bad financial situation, which annoyed me a little. She then starts going on about it being my own fault that I can't afford a dentist, and my own fault that I have a toothache since I haven't been to the dentist for a year or so. (sommarkatten)

The caller experienced feelings of upset over not being understood, and felt that the nurse had engaged in discussing irrelevant matters, in this case her economic responsibilities.
The caller communicated that she was offended. However, another situation was more severe in nature:

I called SHD at about twelve and the nurse there told me that you ‘don’t just feel cold if you have the fever chills’ and that I should call again if I developed pain. Two hours later I wake up and think I’m going to die because of the pain in my left shoulder blade/rib cage up towards the shoulder and neck. There’s no answer at SHD so I call 112 and they send an ambulance immediately. I had SEPSIS and could seriously have died if the pain hadn’t woken me up. (jacktheripper)

An interesting aspect of the scrutinizing game was those callers who adopted a more active role and overrode the nurses’ advice, which in the above example was a matter of life and death. One caller had found that it was sometimes a good idea to call SHD several times, since different nurses answered and would give different, sometimes contradictory, advice. Another way of being “an active patient” was to acquire information on the internet prior to the call to SHD. However, the demonstration of certain medical knowledge could fuel the scrutinizing game further and end up to the caller’s disadvantage:

I’m quite knowledgeable in this area, studying care and nursing at the moment, before I borrowed books from the library, read online and so on for 5 years. So I go straight to the point when I call SHD for example. Tell them my suspicions, explain my symptoms, and finish with the grounds for my concern, for example that I sit all day, hence suspected DVT. But the majority of those who work there have a pretty lousy approach, I think. They have a skeptical tone of voice, kind of hinting that I’m a hypochondriac and just generally a tone that implies ‘Why are you calling and wasting my time’. (lofsan_1979)

Many posters wrote that they used the internet to seek health information before calling SHD, but they did not find that this was appreciated by the nurses at SHD. The poster quoted above had identified the “rules of the game,” and wrote:

But if I call up pretending to be an ignorant person with low general knowledge, I get a much better response […]. (lofsan_1979)

These examples can be interpreted as the manifestation of a scrutinizing game which is a power struggle between the nurse in their role as a professional expert and the patient in their role as an expert on their own body.

Discussion
The objective of the present study was to explore the ways in which callers visiting internet-based forums discuss and communicate their experiences of SHD. The most prominent finding that this study revealed was the scrutinizing game, which included aspects of bidirectional mistrust from both the nurses and the callers. Another important finding was the attitudes that callers held toward nurses who used a technique that was interpreted as “passing the buck.” Hence, this study provides new insights into the theoretical body of knowledge regarding TAN that have not been described in earlier research.

Comparison with prior work
The nurses within TAN hold substantial power, and telenurses have great possibilities to affect healthcare as they steer patient flows within the system. This was also described in the present findings. Communication is essential for the process of telenursing (Greenberg, 2009); but as in most healthcare communication, there is an asymmetry of power between the actors (Fossum, 2007). One problem related to the aspect of power is that it is seen differently by the different actors involved. The holder of the power regards it as obvious and well-intended, but it is perceived as paternalistic and insulting by the receiver of care. However, as previously described, customers also have substantial power, and so discussions in online forums have
the potential to affect attitudes of credibility toward a service (Brown et al., 2007). Negative opinions in particular have been shown to have great impact (Ketelaar et al., 2015). Another way to learn from customer opinions is for healthcare providers to facilitate their clients’ possibility to interact via sponsoring online forums for patients. This kind of consumer group could also be used to highlight and identify possible adverse events that otherwise would remain unknown to healthcare practitioners (Rada, 2008).

In concordance with the present study, and in relation to the conceptualization of cyber nursing, internet forums open up an opportunity to shift power structures (Salzmann-Erikson and Eriksson, 2016; Salzmann-Erikson and Eriksson, 2011). Within healthcare communication, the different actors have different purposes and agendas for the outcome of the call (Fossum, 2007). The nurse’s role as the possessor of something desirable, and the caller’s role as someone in need of help, are both shown within the present results. The findings also demonstrated that callers perceived that the telenurses had a “hidden agenda,” or at least, did not fully inform them of their agenda. This led to callers feeling like they were being diminished and de-intellectualized. Nursing theorists have accounted for the importance of trustful communication styles with patients which involve being open-minded, demonstrating respect, accepting the patient’s story and responding to the patient with honesty and dignity. When nurses do not address these qualities in communication, trust and compliance are threatened (Travelbee, 1971). It is essential that nurses acknowledge a person-centered care and are able to respond differently according to individual problems, needs and stress responses; for example, sick children should be approached in a different way from sick adults, and symptoms of illness may vary greatly between individuals. As previously shown (Ernesater et al., 2012, 2014), telenurses’ communications are “activity-oriented,” using mainly close-ended questions and failing to follow up on callers’ understanding and acceptance of the advice given. This kind of biomedical communication is characterized by the healthcare provider taking control and assuming a leading position in the call. The focus is on the presence or absence of symptoms presented by the patient, and the healthcare provider gives directions but does not follow up the patient’s understanding. Distrust toward the telenurses is an obvious threat to adherence and compliance to advice; and, as shown within the posts in the different forums, provides a basis for questioning the telenurses’ competency. This study shows the existence of bidirectional skepticism between the caller and the telenurse. In their forum posts, callers expressed doubts regarding telenurses’ competence and credibility, as they perceived that the telenurses had mostly found their information on Google. Telenurses, on the other hand, have previously described how they use the information in their computerized decision support system to enhance their credibility with the caller (Ernesater et al., 2009). The issue of distrust has also been expressed by telenurses (Wahlberg et al., 2003); they describe their distrust in callers and their perception that some callers exaggerate their symptoms in order to “get what they want” (Holmstrom and Dall’Alba, 2002). In line with previous research, the present study found that callers perceived telenurses to be questioning their symptoms and the credibility of their narratives. It seems possible that these results are due to telenurses’ strategy of asking questions focusing on excluding symptoms of severe diseases (Holmstrom and Dall’Alba, 2002; Wahlberg et al., 2003), not being aware that this might cause distrust toward them. The issue of telenurses determining and challenging the caller’s reliability as an informant is common, according to a previous review (Kaminsky et al., 2017). The callers were described as both exaggerating and underplaying their symptoms, and telenurses regarded it as important to probe the callers in order to “find the truth” (Holmstrom and Hoglund, 2007; Pettinari and Jessopp, 2001). If callers do not understand the severity of their symptoms, they might find that telenurses use a cautious approach, as shown in the present results. The posts in this study revealed how the callers perceived the telenurses as using “the safe approach,” leaving the caller to decide whether to wait and see
or to seek medical assistance. However, the results also show the well-known problem of the callers distrusting and questioning the telenurses’ advice. The aim of SHD is to increase citizens’ access to healthcare and create a more equal care, but the present results are in line with studies showing that the service tends to be unequal since it is mostly used by demanding callers who are well-versed in the healthcare system (Knowles et al., 2014). Another aim of SHD is to steer patient flows and avoid unnecessary visits to the ER; this was expressed by the callers who perceived SHD as offering a false notion of safety, where the nurses’ actual main task was to keep them away from healthcare providers. In the forum posts, the callers expressed that they expected to have their calls answered immediately. This seems to be a growing problem within today’s healthcare, with care-seekers becoming more demanding and advocating for their patient rights. The results of the study also show how the nurses’ advice and referrals were questioned by other healthcare providers, such as ERs, when the patients arrived. The problem of other healthcare providers not accepting recommendations from TAN has been identified in previous research (Ernesater et al., 2010), but the fact that patients were affected and admonished due to this distrust of telenurses’ recommendations has never been highlighted before.

The conceptualization of cyber nursing as a framework was presented by Salzmann-Erikson and Eriksson (2016) in terms of three aspects: environment, health interactions, and human performativity. In line with this framework, the forum might serve as an environment for an awakening among posters/callers, and hence lead to a rearranged health geography in relation to power structures between the caller and telephone advice nurses. Furthermore, the internet forums are complex in the sense that there is no hierarchical order. When examining how posters/callers interact, it is important to consider the complex mechanisms of torrenting (Salzmann-Erikson and Eriksson, 2011), since they can lead to contagion of attitudes toward TAN. This possible contagion of attitudes should not be underestimated, since cyberspace is not an isolated sphere but rather highly interactive with “real life” decisions (cf. Salzmann-Erikson and Eriksson, 2016). Despite the advantages of the development of internet forums to discuss experiences and attitudes relating to TAN, and the possibility for people to find others with reciprocal experiences, there are also concerns about possible antagonistic effects. The present results imply a potential degradation of trust regarding the general opinion of healthcare services. This might lead to an increased workload at the ER and a higher frequency of calls to the SHD advice line because of repeated calls to obtain a second opinion. Hence, there is a need to continue monitoring virtual discussions about the healthcare system.

Limitations and recommendations for future research
The present study used data from several internet forums. This methodological approach proved to be highly valuable for adding new insights which have not been reported in previous studies. However, it also has limitations. For example, demographic data for the posters are not available. It is also likely that posters who have engaged in forum discussions are not representative of the general population of callers; the sample might represent a population that has a certain interest in the field, which might explain the negative attitudes toward SHD. However, the present sample also included several posts that were both positive and neutral in character. Another limitation that is inherent with the qualitative approach is that we cannot present to what extent the callers are positive or negative toward TAN. Hence, future research can use Sentiment Analysis and Advanced Emotion Analysis which is available in social intelligence platforms (Sanchez-Rada and Iglesias, 2016; Thelwall, 2016).

Practical implications
As shown in the present study, monitoring virtual discussions regarding the healthcare system may provide a new dimension of users’ perceptions of the services, since the anonymity
Bidirectional mistrust

Conclusions
Bidirectional distrust is an extraordinarily important issue which needs to be acknowledged by TAN services since these issues might damage the services on a fundamental level. Trust should therefore be a matter of priority in future development work among leaders and managers. Previous research has pinpointed the inherent limitation of tele-healthcare consultations, namely the absence of bodily communication, which may cause problems of bidirectional credibility/distrust (Holmstrom and Hoglund, 2007; Purc-Stephenson and Thrasher, 2010). However, with existing technology, healthcare providers increasingly offer video calls upon consultations. One way to overcome the bidirectional trust would be to add the dimension of visualization into TAN.

References


Further reading


About the authors

Annica Björkman, RN, PhD, is Senior Lecturer at the Faculty of Health and Occupational Studies, University of Gävle, Sweden, and Department of Public Health and Caring Sciences, Uppsala University, Sweden. Annica Björkman is Registered Nurse Specialized in Primary Care Nursing in addition to holding PhD degree in Medicine and is holding a Position as Senior Lecture at the University of Gävle. Her research focus mainly on nurses work environment, educational research regarding students development and mental illness in primary health care. Annica Björkman is the corresponding author and can be contacted at: anaerr@hig.se

Martin Salzmann-Erikson, RN, PhD, is Associate Professor at the Faculty of Health and Occupational Studies, University of Gävle, Sweden. Martin Salzmann-Erikson is Registered Nurse Specialized in Psychiatric Care in addition to holding a PhD Degree in Medicine with the Specialty in Health and Caring Science and also holding the position as Associate Professor. Dr Salzmann-Erikson has published nearly 60 peer-reviewed papers in international journals and published book chapters for nursing students on basic and advanced level. He has also published popular science articles, contributed to national and international conferences and held lectures and symposia.
Emerald is excited to announce a recent partnership with Peerwith, a platform that provides authors with a variety of services.

The Emerald Peerwith site can be found here: https://authorservices.emeraldpublishing.com/

Peerwith connects academics seeking support for their work with a relevant expert to get their research submission-ready. Peerwith experts can help with the following: language editing, copy editing, scientific editing, translation services, statistical support, funding application support, visuals, video, publication support, literature search, peer review and indexing services. Authors post their assignments on the Peerwith site, experts provide a quote, and the fee and conditions are then agreed upon directly between the author and the expert.

While we are not, of course, guaranteeing publication upon use of Peerwith, we hope that being able to direct academics to this resource either before submission or during the peer review process will help authors further improve the quality of their papers and increase their chances of positive reviews and acceptance.

Academics with relevant expertise can sign up as an expert on the Peerwith system here: https://www.peerwith.com/services/offer
Backfiles Collections

Preserving over 100 years of management research online

A lifetime investment for your institution, Emerald Backfiles will significantly enhance your library’s offering by providing access to over 125,000 articles from more than 260 journals dating back to 1898.

Visit emeraldinsight.com

Get Backfiles Collections for your library

Recommend Backfiles to your librarian today.
Find out more: emeraldpublishing.com/backfilescollections
Internet Research

Number 5
The dark side of social media
Guest Editors: Jari Salo, Matti Mäntymäki and A.K.M. Najmul Islam

1165 Editorial boards
1166 Guest editorial
1169 When social media traumatizes teens: the roles of online risk exposure, coping, and post-traumatic stress
Bridget Christine McHugh, Pamela Wisniewski, Mary Beth Rosson and John M. Carroll
1189 Coping with mobile technology overload in the workplace
Pengzhen Yin, Carol X.J. Ou, Robert M. Davison and Jie Wu
1213 Is my fear of missing out (FOMO) causing fatigue? Advertising, social media fatigue, and the implications for consumers and brands
Laura Frances Bright and Kelty Logan
1228 Investigating microblogging addiction tendency through the lens of uses and gratifications theory
Qian Li, Xunhua Guo, Xue Bai and Wei Xu
1253 Social media’s have-nots: an era of social disenfranchisement
Xinru Page, Pamela Wisniewski, Bart P. Krynenburg and Moses Namara
1275 Tolerating and managing extreme speech on social media
Brett G. Johnson
1292 The dark side of news community forums: opinion manipulation trolls
Todor Mihaylov, Tsvetomila Mihaylova, Preslav Nakov, Lluís Màrquez, Georgi D. Georgiev and Ivan Kolev Koychev
1313 Online moral disengagement and hostile emotions in discussions on hosting immigrants
Francesca D’Errico and Mannella Paciello
1336 The bidirectional mistrust: callers’ online discussions about their experiences of using the national telephone advice service
Annica Björkman and Martin Salzmann-Erikson

ISBN 978-1-78769-903-8

www.emeraldinsight.com/loi/intr