Blackmail on social media: What do we know and what remains unknown?

Al Habsi, A., Butler, M., Percy, A., & Sezer, S. (2021). Blackmail on social media: What do we know and what remains unknown? *Security Journal*, *34*(3), 525-540. https://doi.org/10.1057/s41284-020-00246-2

Published in:
Security Journal

Document Version:
Peer reviewed version

Blackmail on social media: What do we know and what remains unknown?

Abstract

Increasing numbers of people fall victim to blackmail on social media. Yet, there has been little attempt to synthesise research on this topic. This study seeks to address this gap by investigating what is currently known about blackmail and the disclosure of sensitive information on social media. Two rapid reviews were conducted and based on their findings those who disclose more information, are younger, female and willing to use social media to create and distribute images are more likely to fall victim to blackmail on social media. However, worry about privacy and the possibility of becoming a victim of blackmail did not necessarily prevent the disclosure of sensitive information on social media. The implications of these findings for interventions and future research are discussed.

Keywords: cybercrime; social media; blackmail; self-disclosure; social media crime; sensitive information.

Introduction

Incidents of blackmail on social media are increasingly being reported to criminal justice agencies internationally (Al Qahtani, Shehab and Aljohani 2018; Al Salehi 2016; EUROPOL 2017; Office of National Statistics 2019). Yet, official figures are believed to under-represent its true prevalence due to under-reporting (Ahmed et al. 2017). While a substantial body of work exists examining blackmail, less is known about its occurrence on social media. Addressing this gap in knowledge is necessary to inform the design of interventions intended to reduce victimisation and successfully capture and prosecute perpetrators of blackmail on social media. For example, research indicates that people are at risk of becoming a victim of blackmail on social media due to their tendency to disclose sensitive information (Al Qahtani et al. 2018; Al Saggaf 2016; Kopecký 2017; Monaghan 2017). Consequently, understanding why people disclose sensitive information on social media could help with the design of more effective interventions to reduce victimisation. This study seeks to begin to address this gap in knowledge by reviewing existing research on blackmail on social media, as well why people disclose sensitive information on social media.

Blackmail on Social Media

The terms blackmail, extortion and sextortion are regularly used interchangeable, although their meaning varies. Blackmail differs from extortion as extortion involves the use, or threatened use, of force/violence to obtain cash or other valued commodities, while blackmail involves coercion by threatening to expose information about the victim to others (Kopecký 2017). Sancho (2017) argues that online extortion involves threatening victims with the destruction of property or data, while

online blackmail involves coercion by threatening to release sensitive information about the victim that would harm their reputation. Sensitive information has been defined as "describes information that can be used to enable privacy or security harm when placed in the wrong hands" (OHM 2015: 1133). For instance, online extortion could involve criminals hacking a computer and threatening to destroy data. In contrast, online blackmail may involve perpetrators attempting to exert power over victims by threatening to release sensitive information that would harm their reputation, if they do not meet their perpetrator's demands. Sextortion refers to a specific type of online blackmail which focuses on the use of data of a sexual nature to blackmail victims (National Crime Council 2020). On social media, sensitive information can be obtained through the harvesting of social media accounts, user negligence or deliberately and consciously extracting such data from victims.

Blackmail on social media is gradually becoming more common, with information typically being gathered from social media profiles or beguilingly extricated from users themselves (Monaghan 2017). Kopecký (2017) revealed that 6–8 percent of Czech young people (aged 8-17) have experienced blackmail on social media. Al Qahtani and colleagues (2018) have witnessed an increase in blackmail on social media among adults in Saudi Arabia, with blackmailers requesting cash, sex and numerous other items from victims. In Oman, the number of cases of blackmail on social media has also risen, with some victims believed to have committed suicide as a result of their victimisation (Al Salehi 2016). Ahmed and colleagues (2017) similarly report an increase in blackmail on social media in the Gulf Cooperation Council (GCC) countries, with 30,000 such incidents being reported by the GCC per annum and 80 percent of these incidents involving women being targeted with sexual content. The blackmail of children on social media, where erotic evidence or pictures are utilised to force further erotic material, sensual benefits or cash, has also risen steeply in the last decade (EUROPOL 2017). Additionally, companies may be subjected to blackmail on social media as individuals threaten to tarnish their reputation (Raas 2015).

Blackmailers typically request cash, sexual favours or the performance of a legitimate or illegitimate service (Alam 2018; Al Salehi 2016). Being a victim of blackmail can have serious

consequences (Alseyah 2011). It can tarnish a victim's name and family's reputation, as well as lead to psychological distress and even suicide (Al makrami 2015; Monaghan 2017). It can contribute to feelings of anxiety, fear, depression, or social adjustment disorders that may result in social isolation and/or a fear of confronting people (Alseyah 2011). Among children, side effects can include self-blame, invasive memories or feelings, unhappiness, low self-esteem, bad dreams, sleepless, nervousness, anxiety attacks and educational difficulties. Despite these consequences, research indicates that people are often vulnerable to becoming a victim of blackmail on social media due to their tendency to disclose sensitive information online (Al Qahtani et al. 2018; Al Saggaf 2016; Kopecký 2017; Monaghan 2017).

Online Self-Disclosure

Self-disclosure involves sharing sensitive information with other people (Derlaga and Berg 1987). It is a deliberate activity in which a person shares their personal practices, pictures, emotions, etc. with others (Bazarova and Choi 2014). Online self-disclosure has been defined as self-disclosure occurring online (Bauer, Schmid and Strauss 2018). Online self-disclosure is promoted by social media platforms due to its commercial worth (Abramova, Wagner, Krasnova and Buxmann 2017). Currently, the main social media platforms are Facebook, WhatsApp, Twitter, YouTube, LinkedIn, Pinterest, and Instagram (Kapoor et al. 2018). Criminals and members of organised crime groups tend to be more focused on using social media to blackmail users due to the greater tendency of social media users to engage in self-disclosure on social media as compared to other potential targets (Ali et al. 2018; EUROPOL 2017; National Crime Council 2020).

Efforts to combat blackmail on social media have primarily sought to prevent these crimes from occurring by raising public awareness and offering advice on how to avoid becoming a victim of this activity (Al Lawati 2016; Sawer 2016). Frequently, these campaigns seek to discourage people

from sharing sensitive personal information on social media. Nevertheless, people continue to share such information. Some of the reasons put forward to explain this continued involvement in online self-disclosure include a failure to consider the potential costs of sharing sensitive information, instead focusing on the enjoyment they obtain from engaging in this activity and its potential to allow relationships to develop (Bauer and Schiffinger 2016). Governments and criminal justice professionals have struggled to keep pace with social media and the new opportunities it provides for crimes to occur, limiting their ability to reduce victimisation, as well as police and prosecute crimes occurring on social media (Tow, Dell, and Venable 2010; Williams, Butler, Jurek-Loughrey and Sezer 2019; Yar 2018). This study seeks to help to address this gap by synthesising existing research in this area to provide a more holistic understanding of why people may disclosure sensitive information on social media, potentially increasing their vulnerability to becoming a victim of blackmail on social media.

The Present Study

The present study seeks to answer the following two research questions:

Research Question 1: What research has currently been conducted on the occurrence of blackmail on social media?

Research Question 2: Why do people engage in self-disclosure on social media and how may this relate to becoming a victim of blackmail on social media?

By reviewing this literature, it is hoped to identify what is currently known about this behaviour, and what gaps in knowledge remain to be addressed, to enhance our theoretical understanding of this phenomenon and improve our ability to reduce victimisation.

Methodology

Two separate rapid reviews were conducted to answer these research questions. A rapid review is a faster method of synthesising evidence compared to standard systematic reviews (Khangura, Konnyu, Cushmen, Grimshaw and Moher 2012). Rapid reviews provide actionable and relevant evidence in a timely and cost-effective manner, are especially useful in new and emerging areas of research, and give conclusions that do not differ greatly from those of a systematic review (Khangura et al. 2012; National Collaboration Centre for Methods and Tools (NCCMT) 2010; Tricco, Langlois and Straus 2017). Key stages of a rapid review involve developing a search strategy, identifying appropriate databases to search, screening results against agreed inclusion criteria, assessing the quality of the results, extracting key findings from the data and providing a synthesis of the key themes emerging (Campbell et al. 2019).

Search Strategy

Working in conjunction with a specialist librarian, two separate search strategies were developed. For the first rapid review, examining the occurrence of blackmail on social media, the following search terms were used: "(Social media" OR Facebook OR Twitter OR Snapchat OR WhatsApp OR Pinterest OR YouTube OR "social networks" OR "web 2.0" OR webcam OR Internet) AND (blackmail* OR sext*). For the second rapid review, investigating why people engage in online self-disclosure on social media, the following search terms were used: ("Social media" OR Facebook OR Twitter OR Snapchat OR WhatsApp OR Pinterest OR YouTube OR "social network*" OR "web 2.0" OR webcam OR Internet) AND ("self-disclosure" OR "self disclosure" OR "selfdisclosure" Or disclosure OR "self-expression" OR expression OR "self expression" OR "selfexpression" OR "Self presentation") AND (motivat* OR factor* OR reason* OR photo* OR picture* OR Video*). Using these search terms, 18 databases were searched during November 2018. These 18 databases included: International Bibliography of the Social Sciences (IBSS), Sociological Abstracts, PsycInfo, SCOPUS, ProQuest, Social

Services Abstracts, Social Policy and Practice, Social Care Online, QUB PhD theses, ProQuest Dissertations & Theses A&I, EMBASE, Sultan Qaboss University Library Database, IEEE Xplore, Web of Sciences, ScienceDirect, Westlaw Uk, Jil Journal and Google scholar. Conducting a search of these 18 databases using the search terms listed above identified 4043 possible results for the first rapid review and 8612 possible results for the second rapid review. These articles were imported into Mendeley then screened for their eligibility to be included in the rapid reviews, using predefined selection criteria.

Selection Criteria

The articles identified during the database searches needed to meet the following criteria in order to be eligible for inclusion: published from 2004 onwards, peer reviewed, written in English or Arabic, involving primary data collection and examining blackmail or self-disclosure occurring on social media. Only studies published from 2004 were eligible as social media is a recent phenomenon following the establishment of applications like Facebook in 2004, Youtube in 2005, Twitter in 2006, WhatsApp in 2009 and Instagram in 2010. Peer reviewed articles were included to ensure that the studies were of a high standard. Articles written in English or Arabic were eligible for inclusion as these were the languages spoken by the authors and no funds were available for translation. There was a focus on primary data collection to identify what research had been conducted, what this research had found, as well as what gaps in knowledge remained. Lastly, only those studies that focused on blackmail and/or self-disclosure on social media were eligible for inclusion in line with the study's research questions. Articles were excluded from the rapid reviews if they did not meet these criteria or duplicated existing articles. Only 10 articles from the possible 4043 articles identified in the first rapid review were eligible for inclusion, while 67 articles from the possible 8612 articles identified in the second rapid review were eligible for inclusion. Articles were reviewed for their edibility by two members of the research team.

Data Collection and Analysis

The quality of the eligible articles was assessed using the Mixed Methods Appraisal Tool, which has been found to be a reliable tool to use when appraising quantitative, qualitative and mixed research designs (Pace et al. 2012). All articles were deemed to be of sufficient quality for their data to be extracted and analysed. A data extraction tool was developed to extract the data and thematic analysis was used to analyse the data. The findings are presented as a narrative summary, providing an overview of the breadth of research conducted in the area and the main findings that have emerged.

Findings

Based on this analysis, the findings are presented in two sections. The first section reviews the findings emerging from the first rapid review examining blackmail on social media. The second section focuses on the findings emerging from the second rapid review investigating self-disclosure on social media.

Blackmail on Social Media

The first finding to emerge from this rapid review was the scarcity of research that has been conducted on this topic that met the study's inclusion criteria. From an initial 4043 results, only 10 articles met the selection criteria and were eligible for inclusion, despite the growing prevalence of this behaviour in crime statistics internationally. Of these 10 articles, two focused directly on blackmail on social media, with the remainder focusing on social media use more generally or sexual offences occurring online and these articles referencing the occurrence of blackmail in their findings. This

suggests that there is a dearth of research on this topic, which may hinder the development of evidence-based policies, practices and interventions, as policymakers and practitioners do their best to tackle this behaviour in the absence of research evidence specific to blackmail on social media.

Within these 10 articles, three key themes emerged: vulnerability to victimisation; cultural variation in awareness of blackmail on social media; and use of coercive techniques by blackmailers on social media.

Vulnerability to Victimisation

The studies identified several risk factors that could increase vulnerability to blackmail on social media. Age, gender, tendency to engage in online self-disclosure and willingness to use social media to send intimate images were identified as risk factors for becoming a victim of blackmail on social media. Younger people were more at risk of becoming a victim of this activity than older people, with those aged between 15 and 25 years believed to be most at risk (Al Neyadi et al. 2015; Kopecký 2017). Females were also believed to be more at risk than males due to a tendency for females to be put under more pressure to post intimate and sexualised images of themselves on social media (Kopecký 2017; Monaghan 2017). Moreover, a relationship was observed between engaging in self-disclosure on social media and becoming a victim of blackmail on social media, with those disclosing more information being at a greater risk of becoming a victim of blackmail (Kopecký 2016). Similarly, those who were willing to share intimate personal photographs were also at risk of becoming a victim of blackmail on social media (Hamilton-Giachritsis, Hanson, Whittle and Beech 2017; Kopecký 2017; Quayle, Jonsson, Cooper, Traynor and Svedin 2018; Van Ouytsel et al. 2017).

Other factors found to increase a person's vulnerability to becoming a victim of blackmail on social media included a lack of knowledge regarding the legal ramifications of creating and sharing intimate photographs of people under the legal age of sexual consent, as well as the design and usability of social media applications (Hamilton-Giachritsis et al. 2017; Kennedy and Phippen 2018).

For example, young people were often unaware of the legal consequences of creating and sharing intimate images online when under the age of sexual consent, increasingly their likelihood of creating and sharing such images and, consequently, falling victim to blackmail (Kennedy and Phippen 2018). The design and usability of social media applications could also increase vulnerability to blackmail through the promotion of greater self-disclosure, ease of creating and distributing intimate images, permanence of images and level of security and verification surrounding the creation of social media accounts (Hamilton-Giachritsis et al. 2017). Applications whose features encouraged the disclosure of information and images, made the creation and distribution of images easier, kept the images and allowed the creation of accounts with false information could facilitate the occurrence of blackmail.

Cultural Variation in Awareness

Evidence of the important role that cultural norms could play in shaping people's attitudes towards and motivations for engaging in self-disclosure on social media was also apparent, as well as their awareness of the potential to become a victim of blackmail on social media. Studies conducted in Western countries indicated that people often felt that cultural norms encouraged the publication of sexualised content on social media accounts, as people sought to increase their attractiveness and interactions with other users (Al-Makrami 2015; Monaghan 2017; Kennedy and Phippen 2018). In particular, females reported being asked to share intimate photographs, with some stating that they felt pressurised into sharing such images in order to maintain their interactions (Al makrami 2015). Nevertheless, social media users in Western countries reported being less aware of the potential for blackmail or reputational damage to occur as a result of sharing such information on social media (Al makrami 2015). In contrast, cultural norms in Arabic countries tended to discourage the publication of sexualised content on social media profiles (Al makrami 2015; Al Saggaf 2016). This was especially the case for women, as women tended to face greater reputational damage and a reduced social status if they engaged in such behaviour (Al makrami 2015; Al Saggaf 2016). As a result, people living

in Arabic countries, especially women, reported being more aware of the potential for blackmail and reputational damage to occur, if they shared sensitive personal information on social media (Al Saggaf 2016). Indeed, these cultural norms and the potential social consequences that would follow such disclosures led to grave concerns being expressed among Arabic women about their privacy on social media and how the information they shared could be used to blackmail them by other social media users, which was not witnessed to the same extent among females in Western countries (Al makrami 2015).

Yet, despite these fears, cultural norms restricting the development of offline relationships in Arabic cultures, especially for women, were suggested to motivate people to engage in more free and open self-disclosure on social media, as they used social media to develop new relationship, free from the restrictions they faced in their social interactions in the community (Al makrami 2015). Consequently, despite the greater awareness for the potential for blackmail to occur and its serious consequences, people in Arabic cultures continued to place themselves at risk of being blackmailed on social media as they sought to initiate new desirable relationships (Al makrami 2015). Such research raises questions about the extent to which awareness raising campaigns may be successfully in changing behaviour and reducing the risk of blackmail on social media, as well as demonstrate the need for further research to understand how cultural norms, attitudes and beliefs interact with one another to influence behaviour.

Use of Coercive Techniques

The coercive techniques that were used to blackmail both young people and adults on social media were discussed in several articles (Hamilton-Giachritsis et al. 2017; Kennedy and Phippen 2018; Kopecký 2016; Monaghan 2017; Van Ouytsel et al. 2016). Two main tactics were employed by blackmailers to obtain sensitive information from their victims. In the first instance, victims explained they were in a relationship (occurring both online and offline) with their blackmailer and had chosen

to share sensitive information/images with the blackmailer and/or that the blackmailer had recorded them while they voluntarily engaged in a sensitive activity via webcam (Hamilton-Giachritsis et al. 2017; Kennedy and Phippen 2018; Kopecký 2016; Monaghan 2017; Van Ouytsel et al. 2016). It was only when these relationships broke down or when victims were no longer willing to share sensitive information, images or videos with the perpetrator that the blackmail began, with victims being coerced into re-initiating the relationship and/or continuing to disclose sensitive information, images or videos (Hamilton-Giachritsis et al. 2017; Kennedy and Phippen 2018; Kopecký 2016; Monaghan 2017; Van Ouytsel et al. 2016). The second tactic involved the blackmailer trying to emotionally manipulate the victim into disclosing sensitive information by claiming to be depressed and/or suicidal and attempting to induce a sense of guilt/obligation within the victim to help them by sharing such data (Hamilton-Giachritsis et al. 2017). The studies found that both adults and young people were the victims of these coercive tactics.

Disturbingly, studies involving young people also suggest that young people who are victims of blackmail on social media are more likely to report using these coercive tactics against others, perpetuating a cycle of abuse (Kennedy and Phippen 2018; Kopecký 2016). However, it is unclear if adults experience a similar victim-to-perpetrator cycle or what mechanisms may cause this transformation.

Self-Disclosure on Social Media

A larger body of work had been conducted investigating self-disclosure on social media. Of the 67 studies reviewed, the majority (41 out of 67) directly examined why people engaged in self-disclosure on social media or the factors that could increase an individual's probability of engaging in this behaviour. The remainder focused on social media use more generally, with self-disclosure being referenced within their findings. Only two articles considered how self-disclosure may be related to becoming a victim of blackmail on social media, with these articles suggesting that increased

engagement in online self-disclosure increased the risk of becoming a victim of blackmail on social media and highlighting how cultural norms could shape attitudes towards and motivations for engaging in self-disclosure on social media.

Reviewing the 67 articles identified three key themes: the need for relationships and self-expression; the disconnect between privacy concerns and engagement in online self-disclosure; and the impact of platform features on self-disclosure.

Need for Relationships and Self-Expression

One of the biggest factors found to influence self-disclosure on social media was the person's motivations for using social media. Many individuals involved in making online self-disclosures reported being motivated to use social media to develop/maintain relationships, resulting in these individuals engaging in online self-disclosure as they sought to deepen their friendships and gain social status, acceptance, support or validation (e.g. Al makram 2015; Krasnova, Spiekermann, Koroleva and Hildebrand, 2010). Often, these individuals thought about their disclosures and sought to manage the impression they gave other users (Bronstein 2014). These individuals were mostly careful to only disclose information that they believed would help them achieve their relationship goals (Bronstein 2014). In contrast, those who were motivated to use social media for the purpose of self-expression tended to make more spontaneous disclosures as they sought to express themselves online (Seidman 2014). In particular, those who lived in cultures that endorsed social norms which restricted offline behaviour tended to use social media as a means of expressing their 'true' selves and engaged in online self-disclosure as a means of meeting their expressive and emotional needs (Al makram 2015).

Like the studies investigating blackmail on social media, several studies also sought to identify the risk factors that increased the probability of people engaging in self-disclosure on social media. These studies found that age, extraversion, self-control and self-efficacy influenced the amount and type of information disclosed on social media. Those who were younger, more extravert, had lower

self-control and greater self-efficacy in their use and manipulation of social media were more likely to engage in disclosure information on social media (Krämer and Winter 2008; Orzech, Moncur, Durrant, James and Collomosse 2017; Yu 2014). Consequently, the motivations underlying social media use, as well as individual characteristics such as age, personality and psychological attributes, could influence the amount and type of self-disclosure that people engaged in on social media.

Disconnect Between Privacy Concerns and Engagement in Self-Disclosure

The studies also revealed that people were concerned about their privacy on social media. Nevertheless, in most studies investigating how privacy concerns may affect the tendency to engage in self-disclosure on social media, these privacy concerns were not found to directly impact on self-disclosure behaviour. In deciding when to disclosure information on social media, people tended to engage in a cost-benefit analysis of the pros and cons they believed were associated with engaging in online self-disclosure (e.g. Hallam and Zanella 2017; Salleh, Hussein, Mohamed and Aditiawarman 2013). This analysis then informed their decision-making about whether to disclose information on social media or not. In most cases, the cons associated with online self-disclosure were underestimated, while the potential benefits associated with self-disclosure for relationships and self-expression were overestimated (Cheung, Lee and Chan 2015). However, privacy concerns could influence the type of social media that individuals used and the extent to which individuals used the privacy controls on social media applications (Heravi, Mubarak and Choo 2018). For example, privacy concerns could encourage users to limit the visibility of their social media profiles and/or the restrict their interactions with unknown users (Heravi et al 2018). Yet, despite their privacy concerns, these individuals still disclosed personal information to social media users who could view their profiles and with whom they already interacted (Heravi et al. 2018). In this way, privacy concerns could influence what was publicly visible and with whom users interacted but not necessarily what information they disclosed or how much information they disclosed.

Moreover, rather than privacy concerns influencing the amount of self-disclosure individuals engaged in on social media, the extent to which online self-disclosure was a socially accepted norm amongst their peers appeared to play an important role (Van Gool, Van Ouytsel, Ponnet and Walrave 2015). If the making of self-disclosures was a socially accepted norm amongst the users' peers, then these social media users were more inclined to engage in this activity (Van Gool et al. 2015). These studies raise concerns about the potential effectiveness of awareness raising interventions to reduce victimisation. If these interventions are focused on increasing participants privacy concerns as a means of reducing their potential victimisation, then these interventions may meet with limited successful due to the disconnect between privacy concerns and engagement in online self-disclosure witnessed in these studies. Instead, interventions may have more success by tackling the social norms accepting and/or actively encouraging self-disclosure on social media.

Impact of Platform Features on Self-Disclosure

The final key theme to emerge from these studies was the impact that the platform features of social media applications could have on the tendency to engage in self-disclosure on social media. The amount and nature of privacy controls available on different social media applications were found to engender trust (Bevan-Dye and Akpojivi 2016; Krasnova et al. 2010). The more trust users felt in these applications, the more inclined they were to make online self-disclosures (Bevan-Dye and Akpojivi 2016; Malik, Hiekkanen, Dhir and Nieminen 2016; Salleh et al. 2013). Additionally, the extent to which individuals felt that they could trust other users of social media applications influenced their self-disclosure behaviour, with greater feelings of trust linked to more self-disclosures (Chen, Shao and Zhi 2018). The extent to which content could be encrypted, the ease with which images could be taken and shared with others and whether content was temporary or permanent in nature also influenced the extent of self-disclosure on social media. If the content was encrypted, it allowed users to feel in control over who they were sharing their information with and could result in more online

self-disclosure (Ampong, Mensah, Adu, Addae, Omoregie and Ofori 2018; Krasnova et al. 2010). This was despite the possibility of other users sharing this content without their consent. Likewise, if social media applications were easy to use, free and facilitated the sharing of images, these design features tended to encourage greater self-disclosure and image sharing (Bazarova and Choi 2014). Additionally, if individuals believed that their images/data was only going to be shared with others temporarily (e.g. Snapchat), they tended to be more disinhibited in their online self-disclosures and engage in more disclosures (Hofstetter, Rüppell and John 2017). The different design and platform features of social media applications could also influence what individuals used social media application for and, consequently, the amount and type of online self-disclosure that users engaged in on that application. Finally, if individuals were satisfied with the design and platform features of social media applications, they were also more likely to make self-disclosures (Li-Barber 2012). In this way, it is necessary to consider how the design and platform features of social media applications may differentially impact on self-disclosure and, as a result, the potential vulnerability of social media users to becoming a victim of blackmail on social media may vary depending on the social media application they are using.

Discussion and Conclusion

To answer the study's first research question then, very little research has been conducted on the occurrence of blackmail on social media, despite its growing prevalence in crime statistics internationally. This lack of research is important as it can hinder the development of evidence-based polices, practices and interventions, as well as our theoretical understanding of this behaviour. For instance, while criminals and organised crime groups may play a role in victimising individuals by pretending to be in online relationships with victims and then blackmailing them, a focus solely on criminals and organised crime groups will miss how ex-partners of offline relationships can use social media to blackmail victims, as well as how and why young victims of this activity may go on to become perpetrators. This gap in our knowledge, therefore, needs to be addressed to inform the development

of effective interventions, policies and practices in this area, by for example, ensuring we understand how and when young victims of this activity might be at risk of going on to blackmail others so that we can ensure they get the specialist support they need to help prevent this cycle of abuse from occurring and others being victimised.

The second research question sought to understand why some people engage in self-disclosure on social media and how this may relate to becoming a victim of blackmail. In answer to this research question, the findings indicate that those who are seeking a relationship/self-expression, are younger, more extravert, have lower self-control and greater self-efficacy in their use and manipulation of social media will tend to disclose more information online. Those who disclose more information online are at a greater risk of becoming a victim of blackmail on social media, as well as those who are younger, female and willing to use social media to create and distribute images. Interestingly, being concerned about one's privacy and being aware of the potential for blackmail to occur on social media did not necessarily influence the amount of self-disclosure individuals engaged in on social media. Instead, these concerns appeared to influence what was publicly visible on social media profiles and with whom social media users interacted but not what information they disclosed or how much information they disclosed. The design, usability and features of different social media applications could also influence the amount and type of self-disclosure individuals engaged in, resulting in the possibility that some social media applications may place users at a greater risk of blackmail than others.

So, what insights do these findings offer for those working to tackle this behaviour and reduce its occurrence? To begin with, these findings indicate that specialist interventions may be required to support young victims of blackmail on social media to ensure that they do not go on to become perpetrators of this activity. While further research is needed to understand the causal mechanisms underlying any possible victim-to-offender cycle, and if such a cycle can be witnessed among adults, it may be worthwhile exploring if the experience of blackmail on social media may influence what is perceived as 'normal' behaviour on social media and, consequently, how these young people may go

on to interact with other social media users. In addition, these findings raise questions about the potential effectiveness of awareness raising interventions. Most awareness raising interventions seek to reduce victimisation by raising people's awareness of their potential to become a victim of blackmail on social media and/or increase their privacy concerns (Al Lawati 2016; Sawer 2016). Yet, the findings from this research suggest that such actions do not necessarily influence an individual's willingness to disclose sensitive information or what information they disclose, as their need to develop relationships or express themselves can override any caution they may experience as a result of privacy concerns or worries about becoming a victim of blackmail (e.g. Al makram 2015; Heravi et al 2018; Krasnova et al. 2010).

Based on these studies, interventions seeking to reduce victimisation should attempt to address people's underlying motivation for using social media. In other words, interventions should seek to address users desire for self-expression, relationships, social status, acceptance, support and/or validation, and seek to inform users of how they can go about achieving these objectives safely, without putting themselves at risk of blackmail. The wider social norms within society surrounding the acceptability of sharing information on social media, as well as the creation and distribution of sexualised images, especially for women, should also be addressed to reduce the occurrence of blackmail on social media and the pressure people feel to engage in self-disclosure online. It may also be worthwhile to inform social media users of the coercive tactics used by blackmailers to help users recognise when these tactics are being used and help inoculate them against their use. Finally, as the design, usability and features of social media applications can influence self-disclosure and, as a result, vulnerability to blackmail, social media companies may require legislation to encourage them to take greater care in the development of platform features, how these features are used, the security and verification processes surrounding the establishment of social media accounts and to play a more pro-active role in policing the content that is created and shared on their platforms. Through these actions, the potential to become a victim of blackmail on social media may be lessened by reducing the ease

with which sensitive information can be shared, fake accounts created and self-disclosure on social media promoted and encouraged.

Nonetheless, despite the insights offered by reviewing existing research on blackmail on social media and self-disclosure on social media, there are several limitations to this research that must be borne in mind. For example, those articles which did not meet the study's inclusion criteria were excluded from this research, with the result that studies published in languages other than English or Arabic were not examined. Similarly, the focus on published, peer reviewed studies may result in some of the findings being distorted by publication bias, as research which is not novel or significant tends to go unpublished. Accordingly, the exclusion of research which did not meet the study's selection criteria, may limit the potential generalisability of the findings and distort the conclusions that can be drawn from these studies. Future research should seek to address these limitations.

Moreover, synthesising the existing research that has conducted on this topic has identified several important gaps in our knowledge that remain unanswered. For instance, few studies have investigated how differing motivations for using social media may influence the risk of becoming a victim of blackmail on social media. As a result, we do not currently know if those who are motivated to use social media for the purposes of developing and/or maintaining relationships are more likely to become a victim of blackmail on social media compared to those who use social media for self-expression or other purposes. Similarly, whether the experience and prevalence of being blackmailed on social media may vary depending on the type of social media application used has not been investigated. Given that the design and features of social media applications could influence the amount and type of online self-disclosure engaged in, and what users tended to use that social media application for, it is possible that the potential risk of becoming a victim of blackmail may vary across different social media applications. While making users aware of the potential for blackmail to occur on social media applications may not help to reduce self-disclosure, given the disconnect witnessed in existing studies between fear of victimisation, privacy concerns and self-disclosure, such information may nevertheless influence the trust users experience in social media applications and reduce self-

disclosure as a result of the loss of trust experienced, rather than raising awareness of blackmail or privacy fears. Furthermore, how cultural norms and beliefs about the pros and cons associated with disclosing information on social media may affect not only self-disclosure but the risk of becoming a victim of blackmail on social media would benefit from further research. More research is also required on the victim-to-offender cycle referred to in existing studies, to better understand the causal mechanisms that might underlie its occurrence and its possible relevance to adults. There are therefore significant gaps in our knowledge that remain unanswered, hindering our theoretical understanding of this phenomenon and our ability to develop evidence-based policies, practices and interventions on this topic. Future research should seek to address these gaps in knowledge.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Abramova, O., A. Wagner, H. Krasnova, and P. Buxmann. 2017. Understanding self-disclosure on social networking sites-A literature review. *Americas Conference on Information Systems 23*: 1-10.

Ahmed, A., E. Awad, S. Farouq, R. Ayesh, M. Mohammed, N. Alneqabi, M. Abdulahamiad, and J. Ismail. 2017 Online blackmail. Crimes feed on the boom of social networking. https://www.albayan.ae/across-the-uae/news-and-reports/2017-07-24-1.3009066. Accessed 17 December 2019.

Al Lawati, H. 2016. Big jump in cyber blackmail cases in Oman. Retrieved from https://timesofoman.com/article/93557 Accessed 10 February 2020.

Al makrami, H. 2015. *Online self-disclosure across cultures: A study of Facebook use in Saudi Arabia and Australia.* PhD Thesis, Queensland University of Technology, Australia.

Al Qahtani, E., M. Shehab, and A. Aljohani. 2018. The effectiveness of fear appeals in increasing smartphone locking behavior among Saudi Arabians. https://www.usenix.org/conference/soups2018/presentation/qahtani. Accessed 17 December 2019.

Al Saggaf, Y. 2016. An exploratory study of attitudes towards privacy in social media and the threat of blackmail: The views of a group of Saudi women. *The Electronic Journal of Information Systems in Developing Countries 75*(1): 1-16.

Al Salehi, B. 2016. Blackmailing reportage for year 2016, Information Technology Authority. Oman. https://ita.gov.om/ITAPortal/MediaCenter/Document_detail.aspx?NID=109. Accessed 17 December 2019.

Alam, S. 2018. E-extortion: how do teenagers' suicide? https://www.ida2at.com/cyber-extortion-how-do-teenagers-suicide/. Accessed 17 December 2019.

Ali, S., N. Islam, A. Rauf, I. Din, M. Guizani, and J. Rodrigues. 2018. Privacy and Security Issues in Online Social Networks. *Future Internet 10*(114): 1-12.

Alseyah, A. 2011. Symposium on extortion: concept - causes – treatment. Centre for Women Studies. Riyadh. http://www.feqhup.com/uploads/145612377634481.pdf. Accessed 17 December 2019.

Al Neyadi, A., A. Al Kaabi, L. Al Kaabi, M. Al Ghufli, M. Al Shamsi and M. Khan. 2015. Internet Governance and Cyber Crimes In UAE. *International Journal of Scientific & Technology Research*, 4(11): 350-357.

Ampong, G., A. Mensah, A. Adu, J. Addae, O. Omoregie and K. Ofori. 2018. Examining Self-Disclosure on Social Networking Sites: A Flow Theory and Privacy Perspective. *Behavioural Sciences* 8(6): 58.

Bauer, C. and M. Schiffinger. 2016. Perceived Risks and Benefits of Online Self-Disclosure: Affected by Culture? A Meta-Analysis of Cultural Differences as Moderators of Privacy Calculus in Person-to-Crowd Settings. http://aisel.aisnet.org/ecis2016_rp/68 Accessed 10 February 2020.

Bauer, C., K.S. Schmid, and C. Strauss. 2018. An Open Model for Researching the Role of Culture in Online Self-Disclosure. https://aisel.aisnet.org/hicss-51/in/global_issues/3/. Accessed 17 December 2019.

Bazarova, N. and H. Choi. 2014. Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites. *Journal of Communication 64*(4): 635-657.

Bevan-Dye, A. L. and U. Akpojivi. 2016. South African Generation y students' self-disclosure on Facebook. *South African Journal of Psychology* 46(1): 114-129.

Bronstein, J. 2014. Creating possible selves: Information disclosure behaviour on social networks. *Information Research: An International Electronic Journal* 19(1): n1.

Campbell, A., T. Forbes, A. McLaughlin, G. Davidson, M. Butler, C. Blair, N. Menabney and C. McKeaveney. 2019 *Rapid evidence review: The relationship between alcohol and mental health problems.* London: Alcohol Change UK.

Chen, S., B. Shao and K. Zhi. 2018. Predictors of Chinese Users' Location Disclosure Behavior: An Empirical Study on WeChat. *Information* 9(9): 219.

Cheung, C., Z.W. Lee and T.K. Chan. 2015. Self-disclosure in social networking sites: the role of perceived cost, perceived benefits and social influence. *Internet Research 25*(2): 279-299.

Derlaga, V.J. and J.H. Berg. 1987. *Self-disclosure: Theory, research, and therapy*. London: Plenum Press.

EUROPOL. 2017. Online sexual coercion and extortion as a form of crime affecting children: law enforcement perspective. https://www.europol.europa.eu/publications-documents/online-sexual-coercion-and-extortion-form-of-crime-affecting-children-law-enforcement-perspective. Accessed 17 December 2019.

Hallam, C. and G. Zanella. 2017. Online self-disclosure: The privacy paradox explained as a temporally discounted balance between concerns and rewards. *Computers in Human Behaviour* 68: 217-227.

Hamilton-Giachritsis, C., E. Hanson, H.C. Whittle, and A.R. Beech. 2017. *Everyone deserves to be happy and safe. A mixed methods study exploring how online and offline child sexual abuse impact young people and how professionals respond to it.* London: NSPCC.

Heravi, A., S. Mubarak and K.K.R Choo. 2018. Information privacy in online social networks: Uses and gratification perspective. *Computers in Human Behavior 84:* 441-459.

Hofstetter, R., R. Rüppell and L.K. John. 2017. Temporary sharing prompts unrestrained disclosures that leave lasting negative impressions. *Proceedings of the National Academy of Sciences* 114(45): 11902-11907.

Kapoor, K.K., K. Tamilmani, N.P. Rana, P. Patil, Y.K. Dwivedi, and S. Nerur. 2018. Advances in social media research: past, present and future. *Information Systems Frontiers 20*(3): 531-558.

Kennedy, C. and A. Phippen. 2018. Sexting and sexting behaviour-"Oh you're all children, children do silly things. You'll be fine. Get over it!". *Entertainment Law Review,* 28(6): 1-13.

Khangura, S., K. Konnyu, R. Cushman, J. Grimshaw, and D. Moher. 2012. Evidence summaries: the evolution of a rapid review approach. *Systematic reviews 1*(1): 10.

Kopecký, K. 2017. Online blackmail of Czech children focused on so-called “sextortion” (analysis of culprit and victim Behaviours). *Telematics and Informatics 34*(1): 11-19.

Kopecký, K. 2016. Czech Children and Facebook–A quantitative survey. *Telematics and Informatics*, 33(4): 950-958.

Krämer, N. C. and S. Winter. 2008. Impression Management 2.0: The Relationship of Self-Esteem, Extraversion, Self-Efficacy, and Self-Presentation Within Social Networking Sites. *Journal of Media Psychology* 20(3): 106-116.

Krasnova, H., S. Spiekermann, K. Koroleva and T. Hildebrand. 2010. Online social networks: Why we disclose. *Journal of Information Technology 25*(2): 109-125.

Li-Barber, K. T. 2012. Self-disclosure and student satisfaction with Facebook. *Computers in Human Behaviour* 28(2): 624-630.

Malik, A., K. Hiekkanen, A. Dhir and M. Nieminen. 2016. Impact of privacy, trust and user activity on intentions to share Facebook photos. *Journal of Information, Communication and Ethics in Society* 14(4): 364-382.

Monaghan, A. 2017. The impact of Self-Generated Images in online pornography. Middlesex University: Doctoral Dissertation.

National Crime Council. 2020. Sextortion. https://www.nationalcrimeagency.gov.uk/what-we-do/crime-threats/kidnap-and-extortion/sextortion-webcam-blackmail Accessed 27 February 2020.

NCCMT. 2010. *Methods: Synthesis 1. Rapid reviews: Methods and implications*. Hamilton: National Collaborating Centre for Methods and Tools.

Office of National Statistics. 2019. Crime in England and Wales: year ending December 2018. https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingdecember2018 Accessed 27 February 2020.

Ohm, P. 2015. Sensitive Information. Southern California Law Review 88(5): 1125-1196.

Orzech, K. M., W. Moncur, A. Durrant, S. James and J. Collomosse. 2017. Digital photographic practices as expressions of personhood and identity: variations across school leavers and recent retirees. *Visual Studies* 32(4): 313-328.

Pace, R., P. Pluye, G. Bartlett, A.C. Macaulay, J. Salsberg, J. Jagosh, and R. Seller. 2012. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies 49*(1): 47-53.

Quayle, E., L.S. Jonsson, K. Cooper, J. Traynor and C.G. Svedin. 2018. Children in Identified Sexual Images–Who Are they? Self-and Non-Self-Taken Images in the International Child Sexual Exploitation Image Database 2006–2015. *Child Abuse Review*, 27(3): 223-238.

Raas, K.M.R. 2015. *The threat of social media blackmailing in the hospitality industry: when customers misuse their power*. Bachelor's thesis: University of Twente.

Salleh, N., R. Hussein, N. Mohamed and I.J. Aditiawarman. 2013. An empirical study of the factors influencing information disclosure behaviour in social networking sites. In 2013 International Conference on Advanced Computer Science Applications and Technologies (pp. 181-185). IEEE.

Sancho, D. 2017. Digital extortion: A forward-looking view. https://documents.trendmicro.com/assets/wp-digital-extortion-a-forward-looking-view.pdf Accessed 20 February 2020.

Sawer, P. 2016. Huge rise in 'sextortion' by crime gangs using social media to entrap victims. https://www.telegraph.co.uk/news/2016/11/30/huge-rise-sextortion-crime-gangs-using-social-media-entrap-victims/ Accessed 27 February 2020.

Seidman, G. 2014. Expressing the “true self” on Facebook. *Computers in Human Behaviour 31:* 367-372.

Tait, S. E., and D. Jeske. 2015. Hello Stranger! Trust and Self-Disclosure Effects on Online Information Sharing. *International Journal of Cyber behaviour, Psychology and Learning*, 5(1): 42-55.

Tow, W.N.F.H., P. Dell, and J. Venable. 2010. Understanding information disclosure behaviour in Australian Facebook users. *Journal of Information Technology 25*(2): 126-136.

Tricco, A.C., E.V. Langlois, and S.E. Straus. 2017. *Rapid reviews to strengthen health policy and systems: A practical guide*. Geneva: World Health Organisation.

Van Gool, E., J. Van Ouytsel, K. Ponnet and M. Walrave. 2015. To share or not to share? Adolescents’ self-disclosure about peer relationships on Facebook: An application of the Prototype Willingness Model. *Computers in Human Behavior* 44: 230-239.

Van Ouytsel, J., E. Van Gool, M. Walrave, K. Ponnet, and E Peters. 2017. Sexting: adolescents’ perceptions of the applications used for, motives for, and consequences of sexting. *Journal of Youth Studies*, 20(4): 446-470.

Williams, M., M. Butler, A. Jurek-Loughrey, and S. Sezer. 2019. Offensive communications: exploring the challenges involved in policing social media. *Contemporary Social Science*. https://doi.org/10.1080/21582041.2018.1563305.

Yar, M. 2018. A failure to regulate? The demands and dilemmas of tackling illegal content and behaviour on social media. *International Journal of Cybersecurity Intelligence & Cybercrime 1*(1): 5-20.

Yu, S. 2014. Does low self-control explain voluntary disclosure of personal information on the Internet. *Computers in Human Behavior* 37: 210-215.

Word count: 7,588

Date: 12 March 2020.