Abstract | Illegal distribution of sexual images by adults and minors is an expanding problem. We examined whether messages would dissuade males (18-32 years) from visiting a fake website offering access to free pornography to users who uploaded a sexual image of a female. Participants (n=528) seeking the site were randomly assigned to one of three conditions. Group 1 went straight to the landing page. Group 2 encountered a text warning that sharing sexual images of people who appear under 18 years old can be illegal; Group 3 received the same text message with accompanying animation. We measured the attempted click through to the site. Sixty percent of Group 1 participants attempted to access the site compared with 43% in Group 2 and 38% in Group 3. We argue that online messages offer a valuable strategy that can help reduce image-based abuse and the distribution of child sexual abuse material, including by minors.

Warning Messages to Prevent Illegal Sharing of Sexualised Images: Results of a Randomised Controlled Experiment

Prichard, J., Scanlan, J., Krone, T., Spiranovic, C., Watters, P. & Wortley, R.

The consensual sharing of digital sexual images is now a common form of sexual expression among adults (Mori et al., 2020) and adolescents (Madigan et al., 2018). Where these behaviours become exploitative and potentially illegal we find two overlapping constructs of cybercrime: image-based abuse (IBA) and child sexual abuse material (CSAM). IBA research is interested in the abuse of people of all ages – adults and minors – and findings demonstrate unambiguously that IBA occurs in both populations. A study of 4,274 Australians aged 16 to 49 years found that 20% had sexual or nude images taken of themselves without their consent, 11% reported that such images were distributed without their consent and 9% had received a threat that such an image would be distributed (Powell, Henry & Flynn, 2018). This same survey indicated that sharing sexual self-images voluntarily (43%) and/or when unwanted or pressured to do so (41%) were common experiences for young people aged 16-19 years (Henry et al., 2019: 10). Similarly, self-report survey studies have indicated that 38% of Australian youths aged 13-15 year have sent an image of a sexual nature of themselves to others (Lee et al., 2015); 6% aged under 18 years have sent sexualised image of another person without that person's consent (Crofts et al., 2015); and 15% of women aged 15-17 years have experienced IBA (e-Safety, 2017a).

IBA may also constitute CSAM where the person depicted in an image is, or appears to be, a minor. The IBA-CSAM overlap might occur where, for example, an 18-year-old adult photographs their 16-year-old sexual partner. However, the production and distribution of CSAM by offenders who are themselves minors is a rapidly expanding problem (see Dodge & Spencer, 2018; Falligant, Alexander & Burkhart, 2017; Lewis, 2018; McNeish & Scott, 2018). CSAM can be generated by minors in different contexts (see e-Safety, 2017a, 2017b). These include – but are not limited to – the filming of sexual assaults (e.g., Robinson, 2006), non-consensual filming during consensual sex, covert image production (e.g., 'up-skirting'), and 'selfies' of solo sex acts, nudity or sexualised posing (Wolak et al., 2017). Some self-generated material can be produced and consensually shared with one individual (e.g., a boyfriend) which is then non-consensually shared with others. In other situations, minors can be coerced or blackmailed to supply self-generated CSAM to another person.

Responding to these behaviours is a difficult task for government and non-government agencies. Determining the boundary between criminal and non-criminal behaviour is not always simple. Despite being sexual in nature, material generated or shared by minors might not constitute CSAM at law (Lee et al., 2013 p. 34-36). Nor indeed is such activity necessarily harmful (see Lee et al., 2015). There may also be an age-related bar to prosecution, such as instances where children under the age of criminal responsibility generate or share CSAM. These situations may require responses from health and welfare agencies and schools. Where an offence can be proven, there is a broad range of potential offending behaviour – ranging from minor offences through to egregious acts of sexual cruelty – requiring nuanced justice system responses.

The generation or sharing of CSAM by minors can result in significant psychological distress for the young people depicted in the material and their families (e-Safety, 2017b; regarding adult IBA victims see Gassó et al., 2021). A compounding factor is that the material can be uploaded to open websites – 'revenge porn' sites, Facebook, Snapchat, Reddit, Twitter, Tumblr and so forth (Belton & Hollis, 2016; e-Safety, 2017a). On the open Internet, minor-generated CSAM can be accessed by members of paedophilic subcultures, stimulating the CSAM market more broadly. By way of example, of 153 offenders investigated by the Australian Federal Police for CSAM offending, 7.4% possessed CSAM images that were *produced in a school setting* (Krone et al., 2017: 46).

While we have good estimates of the prevalence of IBA, the amount of CSAM produced or shared by Australian minors is difficult to quantify. However, the scale of the problem is potentially very large. At the last census, the nation had approximately 880,000 people aged between 12 and 17 years (ABS, 2016a). Teen-aged Australians had the highest reported rate of Internet access than any other age group in 2016-17 (ABS, 2016b). A range of factors have been identified that put young people at an increased risk of offending online. However, it is the opportunities to offend provided by modern technology – namely common access to digital cameras and the ease of uploading of material to the Internet – that is widely regarded as one of most salient factors (Quayle & Koukopoulos, 2019; Wortley & Smallbone, 2012).

Warning messages as a prevention strategy

While IBA and CSAM are inherently Internet-based phenomena, it is critical to recognise the relevance of scientific research that has examined how people respond to messages in the real 'off-line' world. This research, now many decades old, shows that 'hard-copy' messages can work to mitigate hazards in everyday life arising from motor vehicles, industrial machinery, poisons, pharmaceuticals, foodstuffs and so on.

Importantly, this research has repeatedly demonstrated that the effectiveness of messages is greatly influenced by how they are designed. Table 1 lists design features that increase the likelihood of compliance.

Table 1: Features of message-design that influence human decision-making
Messages are more likely to influence decision-making when they:
Attract attention
Are clear and concise
Impart explicit information about specific hazards, potential harms, and what to do to avoid harm
Are believable
Come from a credible source
Match the degree of danger to specific colours
Match the degree of danger to alert symbols like "!"
Match the degree of danger to signal words like "caution" or "warning"

See Haddad et al., 2020; Prichard et al., 2021.

This evidence base has largely been ignored by crime prevention literature to date. Nonetheless, studies have still demonstrated that messages can work to varying degrees. For instance, postal letters have been shown to be potentially useful to reduce insurance fraud (Blais & Bacher, 2007) and tax evasion (Coleman, 2007) and to protect victims of online fraud (Cross, 2016).

Regarding online warning messages, several studies indicate that Internet users are prepared to heed warnings about hazardous online behaviours related to: perpetrating cyber-attacks (Testa et al., 2017); piracy (Ullman & Silver, 2018); exposure to malware (Haddad et al., 2020); online gambling (Caillon et al., 2020; Gainsbury et al., 2015); pro-anorexia websites (Martijn et al., 2009); and disclosing personal information (Carpenter et al., 2014). While more research is needed to understand what individual factors might increase or decrease the likelihood of compliance with online messages, it is relevant to note findings from Zaikina-Montgomery's (2011) study. She found that warning messages could dissuade users from viewing *legal* pornography. However, an age difference was observed with respect to the warning that "police may be called" – which adolescent participants rated as less effective than did adult participants. One explanation offered for this result was that adolescents were more tech-savvy and consequently less inclined to believe that warning (Zaikina-Montgomery, 2011: 228).

Messages to reduce the viewing of CSAM

Increasing attention has been given to automated online messages as a means of delivering information about CSAM in a targeted fashion to Internet users who have either (a) just commenced offending, or who (b) might be at risk of commencing offending (Prichard et al., 2019). Recommendations to investigate online messages have drawn on theoretical models found in health (Quayle & Koukopoulos, 2019) and crime science (Wortley & Smallbone, 2012).

Law enforcement agencies have trialed CSAM warnings. They are also used by Internet companies (e.g., Google, 2020; see further Prichard et al., 2021). To date the main focus has been on messaging users to dissuade users from viewing, accessing or downloading CSAM. We recently established that online messages can dissuade users from viewing material that eroticises adult-minor sex acts (Prichard et al., 2021). This experiment was a precursor to the study presented below, so the following details are relevant to note.

We subcontracted a commercial agency to design a men's fitness website targeting young adult men, which we refer to as '*GetFit*'. Like other 'honeypot' studies (e.g., Testa et al., 2017), we developed this website to covertly observe the behaviour of anonymous Internet users. Among real

advertisements, we included a fake advert for 'barely legal' pornography, which we used as a proxy for CSAM for legal and ethical reasons (see Prichard et al., 2021: 9). Users who clicked on this advert were randomly allocated to a control group, who received no message, or experimental groups who received different types of warning messages. The messages pretended to come from the administrators of the *GetFit* website. The messages were designed in accordance with the literature presented in Table 1 (above). However, it is important to note that the messages in this initial study did not include any images or animation.

Attempts to enter the barely legal website were made by 73% of the control group (n=100). By contrast, site 'enters' were attempted by 50% of the experimental groups who received a message about police monitoring (n=81) or criminal laws (n=99). These differences were statistically significant and meaningful. The findings were interpreted to support situational crime prevention literature, which has argued that:

- Increasing the perceived risks associated with any offence will reduce the likelihood of the offending behviour (Clarke, 2017); and
- Online CSAM warning messages will be effective because they can reach potential offenders at the moment they contemplate offending (Wortley & Smallbone, 2012).

Testing online messages to prevent image *sharing*

The current study utilised the *GetFit* website and was conducted in collaboration with the Office of the eSafety Commissioner. The primary aim was to test whether automated messages can dissuade young adult males from uploading sexualised images of women to pornographic websites. As outlined earlier, from a public policy perspective, we were particularly interested in the problem of image sharing by minors. However, as explained further below (Recruitment, Ethics), experimental research on minors was not feasible, and so our target sample primarily comprised males between 18-32 years old some of whom may, for example, be considering uploading images of their under-aged girlfriend. Nevertheless, we believe that lessons can be learned from this study more generally about the use of warning messages to prevent the sharing of CSAM by both adults and minors.

A secondary aim was to explore whether a message containing animated graphics would be more effective than a message containing static text alone. We hypothesised that the animated message would be more effective on the basis that congruent text and visuals have been found to enhance attention to and recall of warning messages (e.g. see Lochbuehler et al., 2018).

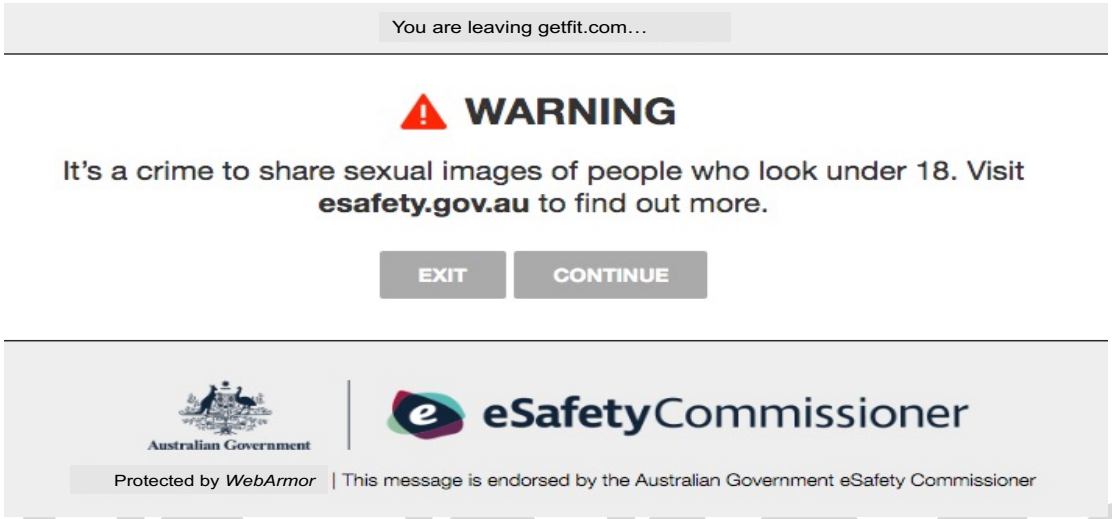
Design and procedure

We conducted a double-blind randomised controlled experiment with naïve participants who visited the *GetFit* website (26/08/2019 – 30/03/2020). For this experiment we presented fake advertisements for a pornography website – referred to in this article as *Swap My Babe* (*SMB*) – which purported to offer users free access to pornography if they uploaded their own sexual image of a woman: "upload your sexiest real pic, free instant access". These advertisements were designed by our commercial partner and simulated advertisements used by real pornography companies.

Users who clicked on the advert were allocated to control or experimental conditions. Allocation was randomized with *Mersenne Twister*. The control group did not receive an online message and therefore could proceed directly to the fake 'landing page' of the *SMB* website. The landing page was also designed by our commercial partner and mimicked landing pages of real pornography sites. It provided users with the option of 'exiting' (navigating to the previous *GetFit* page) or 'entering', which triggered a message after a 5s delay from *SMB*: "Sorry! We're undergoing routine maintenance. Please check back shortly."

Experimental groups were presented with one of two messages. The text presented in both messages stated: “It’s a crime to share sexual images of people who look under 18. Visit esafety.gov.au to find out more”. Message A only contained static text, as depicted below in Figure 1. Message B combined the text with a 9s 2D animation depicting a male character who uploads a sexualised image online and is then arrested.

Figure 1: Visual layout of Message A (text only)



Both messages indicated that the text was endorsed by the Office of the e-Safety Commissioner. Compliance with the evidence base on message-design was achieved by ensuring our messages: were clear, concise and believable; originated from a credible source; and contained an alert symbol (!), and a signal word (‘warning’) (see Table 1). Our messages covered participants’ entire browser screen regardless of the type of device. We also ensured that participants had to interact with the message in order to remove it (e.g. by clicking ‘exit’).

Authenticity and cyber-security

Several steps were taken to ensure that the messages appeared authentic even to tech-savvy participants. This was clearly important for perceived believability, without which the experiment could fail. But it was equally imperative for cyber-security – that is, to reduce the risk of cyber-attacks against the *GetFit* website.

The bottom of the messages contained a small logo for a fictious software package, called here *WebArmour.* Curious users who clicked on this logo were directed to the fake landing page of *WebArmour,* which explained that the software was used by websites to increase cyber-security by scanning, blocking or displaying warnings on outbound links.

As described in greater detail previously (Prichard et al., 2021), we also:

- Used different servers in Australia and overseas to maintain *GetFit*, the *SMB* landing page, the *SMB* advertisements, and the *WebArmour* landing page
- Ensured that *GetFit* and *SMB* were only accessible through SSL, using certificates issued by trusted third parties as opposed to non-SSL sites that lack any proof of authenticity.

Recruitment

The experiment aimed to achieve high ecological validity through the covert observation of the behaviour of users without their knowledge. For legal and ethical reasons we could not recruit individuals under the age of 18 years. Social media advertising was used to attract English-speaking Australian males aged 18-22 years to the *GetFit* website, although Internet users could take other routes to the website (e.g. organic searches). By targeting men only we (a) simplified the design brief for our commercial partner and (b) increased the chance that the limited funds we spent on advertising would achieve a sufficient sample size (since men are more likely than women to use pornography; Rissel et al., 2017).

Outcome measures

Google Analytics provided metrics about the numbers of visitors to *GetFit,* their pathway to the site, and their behaviour at the site. We gathered the IP addresses of the participants who clicked on the *SMB* advert. No other information was gathered about the participants. With respect to our research questions, we measured whether participants attempted to 'enter' the *SMB* website landing page. With this information we created a dichotomous dependent variable, *desistance*.

We deleted repetitions of IP addresses through manual checking and excluded records identified as bots. These procedures mitigated the risk of double counting and ensured that each IP address represented a real individual. Double counting may have occurred if the same participant clicked on the *SMB* advert from different IP addresses (e.g., home and work). On the other hand, since one IP address can be used by multiple users, it is also feasible that we eliminated unique participants from the study.

Ethics

This research was approved by the University of Tasmania human research ethics committee (#H0012409). In accordance with international principles governing human research, despite the fact that the research involved covert observation, the study was conducted for the public benefit and it involved a low risk of causing distress to participants. We employed strategies to protect participants' anonymity, including isolated secure storage of IP addresses (see further Prichard et al., 2021: 8-9). No illegal behaviour was observed. In the graphic design of the *SMB* advertisements and landing page our commercial partner purchased non-pornographic images of certified adult models from a registered company. In all images the models appeared to be adults and appeared to consent to the photography, for example by taking the photograph themselves, or facing the camera.

Results

Metrics provided through Google Analytics indicate that during the seven months that the experiment was conducted *GetFit* was visited from 28,902 unique IP addresses. Most of these IP addresses probably related to single individuals, although we cannot discern how many visitors may have generated more than one IP address (e.g., by visiting the site at work and later at home). Traffic to *GetFit* primarily originated through paid social media advertising (62.4%) and organic searches (i.e., via search engine queries) (32.7%). The remainder came from miscellaneous routes, such as shared links (4.9%).

As detailed below, of the many thousands of visitors to *GetFit* we recorded 528 clicks on the *SMB* advertisements from unique IP addresses after repeat entries and non-human agents (e.g. bots) were excluded. This number represents 1.83% of all *GetFit* visitors, which by industry standards is good click-through rate for web-based display advertisements (e.g. Irvine, 2020). The main weakness of our method is that the demographic profile of the participants is unknown. However, we can

conclude with some confidence that the participants were likely to be Australian males aged between 18 and 32 years. This is for two main reasons.

First, the social media advertisements targeted Australian males aged 18-22 years, which suggests that approximately 18,000 of the 28,902 visitors (i.e. 62.4%) fitted this demographic profile. (We acknowledge though that some individuals may lie about their age when they create their social media accounts.)

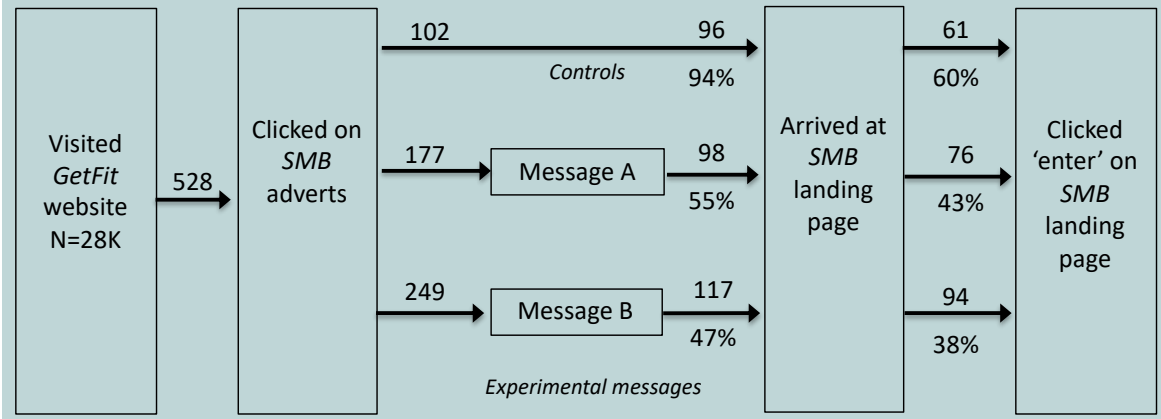
Secondly, our commercial partner advised that the organic traffic (32.7%) was probably mainly made up of the Internet users we attracted to the *GetFit* website in the first experiment (Prichard et al., 2021). The number of visitors in that experiment was estimated to be 29,364. Over 90% of these arrived through our social media advertising campaign, in which we targeted Australian males aged 18-30 years. Taking into account that first experiment started in November 2017 – almost two years before the current experiment – we conclude that this 'original' cohort would now be aged between 20 and 32 years.

Desistance

The behaviour of the 528 participants who clicked on one of the *SMB* advertisements is presented in Figure 2. 'Desistance' refers to the proportion of participants who did not click 'enter' at the *SMB* landing page.

Our cell sizes were uneven due to the randomisation algorithm that we employed (a *Mersenne Twister*): 249 participants in the Message B group compared with 102 in the control. This algorithm ensured that the temporal allocation of a participant to an experimental group or control was independent of all prior allocations, ensuring no sequencing bias. However, the downside of this approach is that it does not guarantee an equal allocation of participants to any particular group. Future studies will utilise an approach that guarantees sequential randomness as well as equal allocation as far as possible. Of the 102 participants in the control group 96 arrived at the *SMB* landing page. Six did not. How this occurred is not clear. This issue is considered further in the Discussion.

Figure 2: Number of participants who arrived at the *SMB* landing page and clicked 'enter'



At the *SMB* landing page 61 members of the control group attempted to gain access to the *SMB* site by clicking 'enter'. This means that the control desistance rate was 40%. The experimental groups

had two opportunities for desistance. Their first opportunity was when they received a warning message which prevented them from advancing to the *SMB* site until they navigated away or confirmed their intention to visit the site. The second opportunity for desistance was – like the controls – at the *SMB* site. Of the Message A participants (N=177), 98 continued to the *SMB* landing page and once there 76 clicked 'enter'. In the Message B group (N=249) 117 continued to the landing page and 94 clicked 'enter'.

Both of the experimental groups were separately compared with the controls. Fisher's Exact Test (1-sided) was applied to determine the statistical significance of the observed proportion of users from each group who did not click 'enter' (i.e., desistance) on the *SMB* landing page. (In contrast to Chi-Square, there is no statistical test value to report when conducting the Fisher's Exact Test.) For Message A and B, respective desistance rates were 57% and 62%. Individual comparisons found that the difference in desistance rate between the control group and Message A [p= .005, OR= 1.977 (95% CI OR = 1.205, 3.244)] and Message B [(p < .001, OR= 2.453 (95% CI OR = 1.531, 3.931)] were statistically significant.

The effect sizes, as indicated by odds ratios (OR), were small but either exceeded (Message B) or approached (Message A) the minimum threshold (i.e., OR = 2.0) for a difference which is practically meaningful according to Ferguson's (2009) guidelines for the social sciences. In other words, the magnitude of the difference in desistance rates for the control group compared with the two message groups was meaningful. A pairwise comparison of the difference in desistence rate between Message A and B was not significant (p = .16) indicating the messages were equally effective.

Discussion

Our study has demonstrated that individuals can be dissuaded from sharing sexualised images if they received an online warning message concerning a potential breach of CSAM laws. We believe that our findings are robust because the participants did not know they were being observed, and by randomly assigning the participants to control or experimental conditions we controlled for selection bias or other factors.

Importantly, it is very likely that participants exhibited real-life behaviours. Every dimension of *GetFit* and *SMB* was professionally designed. In fact, the *SMB* advert was markedly more attractive to users than the advert we used in the original barely legal experiment. This is evidenced by the fact that the original advert took 16 months to recruit a sufficient sample size, whereas the *SMB* advert took seven. Whether this was due to the type of pornography *SMB* purported to offer, or the physical appearance of the adverts is unclear. But either way, the overall 'pitch' was evidently realistic from the perspective of Internet users. Other indications that none of the users suspected the true research purposes of the websites include: the fact that no complaints were received from users by the "contact us" email address at *GetFit*; and neither *GetFit* nor *SMB* were subject to system trespass (hacking) attempts.

The warning messages used in this experiment: used colour and alert symbols appropriately; were simple and believable; did not disappear from participants' screens until they navigated away; and appeared to come from a highly credible source (the Office of the eSafety Commissioner). Given these features, it is perhaps not surprising that animation did not significantly increase message-effectiveness. The animation was redundant because the other features of the message had already caught participants' attention.

Of the many thousands of visitors to our men's fitness website, *GetFit*, 528 chose to click on the professionally designed *SMB* advert, which offered free pornography to users if they uploaded their own images of women. The warning significantly reduced the click-through rate to the *SMB* website. The plainest interpretation of this finding is that – consistent with situational crime prevention literature (Clarke, 2008, 2017; Wortley & Smallbone, 2012) – the warnings influenced some, but not all, participants' decision making by increasing their perception of the risks associated with the *SMB* website.

We believe that the results of our study can be generalised in two ways. First, although the messages used in the experiment specifically referred to the uploading of underage images, it is highly likely that they had a broader effect. While some of the participants may have been actually prepared to upload their own sexual image of a women at the *SMB* website, others might have simply been curious about the pornography *SMB* purported to offer. For such individuals the warning message may have triggered generalised fears about the legality of the content at *SMB.* In other words they may have wondered whether *SMB* contained CSAM. However, even where this was the case, the messages still demonstrated their crime prevention capacity by reducing the *likelihood* of distribution. This is because the messages diverted users away from a potentially criminogenic online environment – an online "situation" where they could decide to share an image, regardless of their intentions prior to arriving at the *SMB* site. In situational crime prevention terminology, the phenomenon whereby an intervention has an effect beyond its specific target is known as diffusion of benefits (e.g., Guerette & Bowers, 2009). Just the same, we might expect that specific warnings to prevent IBA among adult populations would be even more effective if the content of the message referred to the criminality of some IBA behaviours.

Secondly, despite the fact that our participants are likely to have been adult males aged 18-32 years, in our view the findings have implications for the prevention of the sharing of CSAM by minors – particularly those in the 15–17-year-old age bracket. This age group has demonstrated similar patterns of behaviour to adults with regards to IBA (e.g., e-Safety, 2017b; Powell, Henry & Flynn, 2018). Additionally, since the wording of our messages was simple, we do not see any grounds for concerns about aged-based differences in literacy. Nor does there appear to be a sound reason to expect that the types of age differences observed by Zaikina-Montgomery (2011) might operate if our message was trialed with 15–17-year-olds. In Zaikina-Montgomery's (2011) study, unlike the adult participants, adolescent participants appeared disinclined to believe a warning that police would be called after an attempt to access legal pornography. However, we see no reason for our message to stretch credulity in the minds of adolescents because our warning is a simple statement of fact; it *is* an offence to share images of sexualised images of people who look under the age of 18 years in Australia.

Agencies that embark on developing messages specifically to target 15–17-year-olds would need to ensure that they adhered to the literature on message design. In our view, messages could be appropriate and beneficial for younger age-groups, including those under the age of criminal responsibility – particular if messages were used to assist children to increase their online safety and reduce the risks of victimisation. However, very little indeed is known about message design in this context and great care would be warranted.

In terms of limitations of this study, as discussed the main drawback of our honeypot method is the lack of data it yields about our participants. However, as we have argued previously (Prichard et al., 2021), this limitation is a fair trade-off for the benefits we have identified above. Six participants in the control group did not arrive at the *SMB* landing page. Why this occurred is not clear. Slow connectivity or computer speed might have provided a few seconds for the participants to lose interest and navigate away. We cannot discern whether similar factors affected any participants in

the experimental groups. Future experiments on messages that use the honeypot method will need to identify whether additional data can be ethically collected about participants' online behaviour whilst adequately protecting their anonymity.

Funding

This project was funded by the Australian Research Council (DP160100601) and the Australian Institute of Criminology.

References

Australian Bureau of Statistics (ABS) 2016a, *Household use of information technology,* viewed 23 November 2018, http://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0

Australian Bureau of Statistics (ABS) 2016b, *Australian Demographic Statistics,* viewed 23 November 2018, http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Jun%202016

ANSI (American National Standards Institute). 2016. *American National Standard Design Principles for Environmental/Facility Safety Signs and Product Labels* (ANSI Z535.X-2016). National Electrical Manufacturers Association. https://doi.org/10.1371/journal.pone.0082055

Belton, E., & Hollis, V 2016, *A review of the research on children and young people who display harmful sexual behaviour online*, viewed 26 November 2018, https://learning.nspcc.org.uk/research-resources/2016/review-children-young-people-harmful-sexual-behaviour-online/

Blais, E., & Bacher, J. L. (2007). Situational deterrence and claim padding: results from a randomized field experiment. *Journal of experimental criminology*, *3*(4), 337-352.

Caillon, J., Grall-Bronnec, M., Saillard, A., Leboucher, J., Péré, M., & Challet-Bouju, G. (2021). Impact of Warning Pop-Up Messages on the Gambling Behaviour, Craving, and Cognitions of Online Gamblers: A Randomized Controlled Trial. *Frontiers in Psychiatry*, 12, 711431–711431.

Carpenter, S., Shreeves, M., Brown, P., Zhu, F., & Zeng, M. (2018). Designing warnings to reduce identity disclosure. *International Journal of Human–Computer Interaction*, 34(11), 1077–1084.

Clarke, R. V. (2017). Situational crime prevention. In R. Wortley & M. Townsley (Eds.), *Environmental criminology and crime analysis* (2nd ed., pp. 286–303). Routledge.

Coleman, S. (2007) *The Minnesota Income Tax Compliance Experiment: Replication of the Social Norms Experiment*. Available at SSRN: https://ssrn.com/abstract=1393292

Crofts T, Lee M, McGovern A & Milivojevic S 2015. *Sexting and young people*. London, UK: Palgrave Macmillan.

Cross, C. (2016). Using financial intelligence to target online fraud victimisation: applying a tertiary prevention perspective. *Criminal Justice Studies*, *29*(2), 125-142.

Dodge, A., & Spencer, D. C. (2018). Online sexual violence, child pornography or something else entirely? Police responses to non-consensual intimate image sharing among youth. *Social & Legal Studies*, *27*(5), 636-657.

Falligant, J. M., Alexander, A. A., and Burkhart, B. R. 2017, 'Risk assessment of juveniles adjudicated for possession of child sexual exploitation material', *Journal of Forensic Psychology Research and Practice*, vol. 17, no. 2, pp. 145-156.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *5*, 532–538

Gainsbury, S., Aro, D., Ball, D., Tobar, C., & Russell, A. (2015). Determining optimal placement for pop-up messages: Evaluation of a live trial of dynamic warning messages for electronic gaming machines*. International Gambling Studies*, 15(1), 141–158.

Gassó, A. M., Mueller-Johnson, K., & Gómez-Durán, E. L. (2021). Victimization as a Result of Non-Consensual Dissemination of Sexting and Psychopathology Correlates: An Exploratory Analysis. *International Journal of Environmental Research and Public Health*, *18*(12), 6564.

Guerette, R. T., & Bowers, K. J. (2009). Assessing the extent of crime displacement and diffusion of benefits: A review of situational crime prevention evaluations. *Criminology*, 47(4), 1331-1368.

Google. (2020). Fighting child sexual abuse online. https://protectingchildren.google/intl/en/

Grant, H. (2020) World's biggest porn site under fire over rape and abuse videos, *The Guardian,* viewed 21 June 2021, https://www.theguardian.com/global-development/2020/mar/09/worlds-biggest-porn-site-under-fire-over-videos-pornhub

Haddad, A., Sauer, J., Prichard, J., Spiranovic, C., & Gelb, K. (2020). Gaming Tasks as a Method for Studying the Impact of Warning Messages on Information Behaviour. *Library Trends*, *68*(4), 576-598.

Henry, N., Flynn, A. & Powell, A. (2019). Image-based sexual abuse: Victims and perpetrators. *Trends & Issues in Crime and Criminal Justice (No. 572).* Australian Institute of Criminology.

Krone, T., Smith, R.G., Cartwright, J., Hutchings, A., Tomison, A. and Napier, S 2017, *Online child sexual exploitation offenders: A study of Australian law enforcement data*, Report to the Criminology Research Advisory Council, Canberra.

Lewis, R 2018, 'Literature review on children and young people demonstrating technology-assisted harmful sexual behaviour', *Aggression and Violent Behaviour*, vol. 40, pp. 1-11.

Lee, M., Crofts, T., McGovern, A. and Milivojevic, S 2015, 'Sexting among young people: Perceptions and practices', *Trends and Issues in Crime and Criminal Justice*, no. 508, pp.1-9.

Lee, M., Crofts, T., Salter, M., Milivojevic, S., McGovern, A 2013, '"Let's Get Sexting": Risk, Power, Sex and criminalisation in the Moral Domain', *International Journal for Crime and Justice,* vol. 2, no. 1, pp. 35-49.

Lochbuehler, K., Mercincavage, M., Tang, K. Z., Tomlin, C. D., Cappella, J. N., & Strasser, A. A. (2018). Effect of message congruency on attention and recall in pictorial health warning labels. *Tobacco Control*, 27(3), 266-271.

Madigan, S., Ly, A., Rash, C. L., Van Ouytsel, J., & Temple, J. R. (2018). Prevalence of multiple forms of sexting behavior among youth: A systematic review and meta-analysis. *JAMA pediatrics*, *172*(4), 327-335.

Martijn, C., Smeets, E., Jansen, A., Hoeymans, N., & Schoemaker, C. (2009). Don't get the message: The effect of a warning text before visiting a pro-anorexia website. *International Journal of Eating Disorders*, 42(2), 139–145.

Mori, C., Cooke, J. E., Temple, J. R., Anh, L., Lu, Y., Anderson, N., ... & Madigan, S. (2020). The prevalence of sexting behaviors among emerging adults: A meta-analysis. *Archives of sexual behavior*, *49*(4), 1103-1119.

McNeish, D., & Scott, S 2018, *Key messages from research on children and young people who display harmful sexual behaviour*, viewed 26 November 2018, https://www.csacentre.org.uk/research-publications/key-messages/harmful-sexual-behaviour/

Office of the eSafety Commissioner (2017a), *Image-based abuse national survey: summary report,* viewed 21/06/2021, https://www.esafety.gov.au/sites/default/files/2019-07/Image-based-abuse-national-survey-summary-report-2017.pdf

Office of the e-Safety Commissioner (2017b), *Image-based abuse: qualitative research summary*, viewed 21/06/2021, https://www.esafety.gov.au/sites/default/files/2019-07/Image-based-abuse-qualitative-research-summary-2017.pdf

Powell, A., Henry, N., & Flynn, A. (2018). Image-based sexual abuse. *Routledge handbook of critical criminology*, 305-315.

Prichard, J., Wortley, R., Watters, P. A., Spiranovic, C., Hunn, C., & Krone, T. (2021). Effects of Automated Messages on Internet Users Attempting to Access "Barely Legal" Pornography. *Sexual Abuse*, 10790632211013809.

Prichard, J., Krone, T., Spiranovic, C., & Watters, P. (2019). Transdisciplinary research in virtual space: Can online warning messages reduce engagement with child exploitation mate- rial? In R. Wortley, A. Sidebottom, N. Tilley, & G. Laycock (Eds.), *Routledge handbook of crime science* (pp. 309–319). Routledge.

Quayle, E., & Koukopoulos, N. (2019). Deterrence of online child sexual abuse and exploitation. *Policing*, *13*(3), 345–362.

Rissel, C., Richters, J., Visser, R. O. D., McKee, A., Yeung, A., & Caruana, T. (2017). A profile of pornography users in Australia: Findings from the second Australian study of health and relationships. *The Journal of Sex Research*, 2, 227–240.

Robinson, N. (2006). 'Assault DVD suspects may all face charges'. *The Australian.* October 25, 2006. Accessed 18 March 2021. https://web.archive.org/web/20061109020811/http://www.theaustralian.news.com.au/story/0%2C20867%2C20642653-5006785%2C00.html

Testa, A., Maimon, D., Sobesto, B., & Cukier, M. (2017). Illegal roaming and file manipulation on target computers: Assessing the effect of sanction threats on system trespassers' online behaviors. *Criminology and Public Policy*, 16(3), 689–726.

Ullman, J. R., & Silver, N. C. (2018). Perceived effectiveness of potential music piracy warnings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1), 1353–1357.

Wolak, J., Finkelhor, D., Walsh, W. and Treitman, L 2018, 'Sextortion of minors: characteristics and dynamics', *Journal of Adolescent Health*, vol. 62, no. 1, pp. 72-79.

Wortley, R., & Smallbone, S. (2012). *Internet child pornography: Causes, investigation, and prevention*. ABC-CLIO.

Zaikina-Montgomery, H. (2011). *The dilemma of minors' access to adult content on the internet: A proposed warnings solution* [Doctoral thesis]. University of Nevada.