APPG ON CYBER SECURITY MEETING AGENDA 20th April at 4.00 p.m.

Title: The meeting will look at the Online Harms Bill from a Cyber Security perspectives. It will particularly take evidence about what is considered to be missing from the Bill and areas that it could therefore usefully cover.

Chairman's welcome – Christmas tree bill, looked at in twice in Committee. Covers a huge range of topics, many of which are contentious. Some areas to applaud: codifying harmfulness, anti-terrorism etc. Talks about boosting tech sector safety and international consensus.
Present: Simon Fell MP (Chair), Lord Arbuthnot of Edrom, Lord West of Spithead
Apologies: Baroness Neville Rolfe, Lord Mackenzie of Framwellgate
Speakers:

1. **Dr Konstantinos Mersinas is the Director of the Distance Learning MSc Programme in Information Security and the Academic Lead for industrial placements at the Information Security Group, Royal Holloway University of London.**

Three components underlying the Bill:

a. Not all harmful content is illegal, talking about obscene material and indecent images of children, extreme pornography (extreme = participants being in physical danger of injury or death)
b. E-whoring – selling of personal data in a deceptive fashion
c. Cyber stalking, bullying and grooming. Challenge is that some groups are more vulnerable than others, e.g. children so no single solution to fit all issues.

To understand the behaviour of individuals online we need to understand how communications happen online. Then we are need to look at individuals' personality and the online environment.

Social engineering – use of deception to manipulate individuals to reveal information. Number of psychological principles which underly this behaviour shown in slide. These principles are used by bad actors for grooming, radicalization, hate speech and even disinformation. Personality traits – big 5 factor or OCEAN model. We all have aspects of these traits, some are higher or lower than others in our personalities. The Dark Triad is linked to criminal behaviour, different traits manifest different behaviours. Research reveals certain patterns of behaviour and crimes being linked with specific combinations of personality traits; we know, e.g. that Cyber bullies are low on agreeableness and conscientiousness. Some people are more susceptible to the effects of cyber attacks.

3rd component is the environment – the online disinhibition effect, people feel that they are anonymous and invisible, and communication is different to face-to-face, you can send a message and come back to it the following day for instance. The online disinhibition effect causes people to behave offensively, although they would not do so in a face-to-face discussion.

There is however a positive side: e.g. shy people find it easier to express opinions online rather than face to face.
Platform design needs to include behavioural functionality (the so-called nudges and boosts). User's motivations need to be invoked and trigger the user to take action and users' ability to perform a task is important, this can be achieved by making the task, e.g. easy. Another model/approach for nudges and boosts is EAST (easy, attractive, social and timely) which describes the characteristics that behavioural interventions need to have.
We can use AI, but it is limited. Can work *with* the users or *for* them, e.g. we are working on an AI user assistant at RJUL to be used when and if the user decides to do so. Need to understand and take into consideration:

- Influencers
- Individual
- Environment

2. Stewart Room Partner
Global Head of Data Protection & Cyber Security, DWF Law LLP

Four issues to bring out looking at gaps and where else we might look for solutions to gaps.

i – Platforms covered by the White Paper being used as a malware distribution centres or hacker shop fronts. Plenty of empirical evidence about this: malware (including links thereto) is frequently found in platforms on the Surface Web (i.e., this is not simply a Dark Web issue) and these platforms can be used to advertise hacking services and tools. The information immediately underneath the table of harms (which identifies the main areas of harm of concern) in Part 1 of the White Paper, says that all harms suffered directly due to a breach of cybersecurity or hacking, are not covered in the bill. We need to be clear on the meaning of the idea of direct causation: what does this mean exactly? Moreover, the implicit claim that direct cyber security harms are covered by the national security strategy in incorrect. There are clear gaps in the regulatory framework.

For example, when thinking about the idea of the distribution centre and shop front, the CMA has prohibitions against making articles (etc.) available for computer misuse offences, but this is targeting offenders, not platforms. That area of the law can only have a limited impact in comparison on the impact that would be had if the distribution centre and shop front where "closed". The idea of how to stop distribution/ the shop front within the platform is not addressed by the CMA.

For example, take the NIS regulations. Currently, they do not cover most of the platforms that fall within the scope of the White Paper, but even if they did they would not address the distribution centre/ shop front problem, because the NIS regulations are focused on the

underlying security of the NIS that the platforms rely upon to operate: they do not protect individuals from the mechanisms of harm that the White Paper is concerned with.

Therefore, there is a potential regulatory here. We could miss the opportunity to address hacking and malware by taking cyber security issues out of scope on the forthcoming Bill. Even if these points are not addressed in the Bill, the gap is nevertheless worth investigation.

ii – interaction with data protection law – again, see the bullets under Table one, Part One. Again, it is important that we focus on the meaning of the word: "directly". However, even if harms to individuals can be covered by the data protection regime, that is not a reason for them to be excluded from this law. Take site scrapping of personal data as an example. That is a data protection issue, if we want to call it that. However, if we analysed it differently, we might call it a preparatory steps towards the harassment and abuse that Andrea gave evidence on. Thus, the fact that a harm might be directly covered by data protection law is not a cogent reason in itself to exclude the harm from the Bill. This needs re-thinking and care to get right.

iii- encryption and clarity around this. This issue is dealt with in para 33 of the Exec Summary, but the language is oblique. We need to understand more of the intent, because encryption can be used to disguise harms, but it is also a failsafe against other harms and attacks on encryption, while justifiable on an issue by issue basis need rigorous analysis and debate. Also, we need to take a more holistic view of the civil liberties issues, noting the current adjustments to civil liberties due to Covid, which are unprecedented. We need to look at privacy and civil liberties in the round when we are developing surveillance structures in society at an unprecedented pace.

iv – enforcement regime. Once a controller of data falls into operational failure, 3rd parties need to step in to help cure any harms (i.e., regulators and people affected) so a transparency mechanism after operational failure serves a lot of public good. However, those regimes which have hugely important public policy objectives can create perverse incentives to "bury the bad news". For example, the good behaviour of transparency after operational failure creates a conveyor belt for regulatory fines and penalties, compensation claims, litigation and other contentious legal cases. It is not surprising, then, that in the real world, on the ground, the concern is as much about "back covering" as it is about transparency. This leads to distortions in behaviours and outcomes, ultimately to the disadvantage of the policy objective.

The speaker supports breach notification and related transparency ideas, but believes in a step between transparency and punishment so that an organisation that does the right thing after operational failure (by being transparent) is rewarded, rather than pushed into a zone of immediate "punishment" and litigation risk. There is a risk that the financial penalty regime proposed in the White Paper apes and mirrors the GDPR, which will bake in the same risk of

perverse disincentives. The GDPR is a great advance for it does have some fundamental flaws in it, of which this is one. There should there be a "firewall" between transparent and punishment/ liability.

3. Andrea Simon

Director at End Violence Against Women Coalition (EVAW)

EVAW is a UK-wide coalition of more than 100 women's organisations and expert members working to end all forms of violence against women and girls (VAWG) such as domestic abuse and sexual violence, forced marriage, FGM, stalking, sexual harassment and more. Set up in 2005 to address these issues.

Current proposals do not go far enough for women. Want to see more detail about the duty of care that tech companies have, should go beyond the removal of content. Ofcom should be able to hold tech firms accountable for the safe design of their systems.
Slide 2 – Online harms should be considered as part of general abuse which women suffer. Women are 27 times more likely to be harassed than men for instance. Online violence targets women and girls above all. Experience of online abuse has a huge impact on people's behaviour and willingness to use technology.
Slide 3 – what is actually happening to women? Harassment occurs both on and offline, the pandemic has led to an increase in this. Online abuse is not really different to the sort of abuse given in the real world. It contributes to the withdrawal of certain groups from public life.
Slide 4 – shows some of the common experiences of black and minoritized women.
Slide 5 – key issues and recommendations. Keen to see detailed recognition of violence against women and girls in all online forms. Current proposals do not recognize the extent and nature of abuse.
How duty of care is enforced is also key, regulator should have a mandate for safe design of systems.

Open questions and discussion –

Simon Fell – Andrea and Konstantinos mentioned safe design, what are the markers for a robust, safe system?
AS – about the starting point, we find that safety is retro-fitted into the technology. Many platforms have been around for some time. We can look at more ethical frameworks, analysis of harms and have a two-step process.
KM – encouraged to hear the techniques mentioned by AS. We have seen this in our research, there is a disconnect between the perceptions of what is harmful and what is not. Mentioned comments on an author's website, made by children who did it for fun but caused the author to be depressed. In traditional cyber security the human is the final line of defence. Technical or legal measures are not the final one, human intervention is needed in an online platform.

Keith Scott – has a PhD student who is looking at techniques around the destruction of personality as part of abuse. Interested in future proofing, we have seen this in embryo research, how do we legislate for something does not exist? Has no answer, what are the necessary approaches?

AS – concerned that we are playing catch up, particularly in relation to AI. Online Harms bill framing must be forward looking. What can we put in place to recognize advancements in methods of harassment that are flexible and can include other harms?

KS – focus on the end results and the individual not the tech.

Babita Kumar – wanted to ask how does AI become proactive as a defence mechanism?

KM – mentioned the use of AI, work has to do with companies. There is contact between employer and employee. Google and FB understand how to do this, clever move to have them work with the Govt. Senses the obstacles are initially significant as there is a high level of surveillance implied. AI could work with individuals who might be victimized so that they can learn how to become aware of issue and how to defend themselves.

SR – KM is correct that there are complicated questions about how to build profiles and how humans should intervene. Would not want a fully automated decision process. The Bridges / South Wales case about automatic face recognition case is interesting. The decision was based not on a substantive privacy effect but one of policy and paperwork. The AFR trial was unlawful because the policy framework was obtuse and data protection was weak. If the police had been more careful around policy and risk assessment then the Court would have tolerated the scanning of the public. Use of AI is the same, the Judiciary is with us on this point and against the failure of policy and paperwork.

Lord West – concerned that there is a real risk when you start getting too much data. Mass facial recognition is not liked by the Judiciary, many get worried about Big Brother behaviour.

SR – the cohort of concerned people included many with an interest in civil rights. There is a huge amount of concern. If you can demonstrate that the surveillance activity has a negligible effect the Judiciary would allow it to happen. The actor is also important, as well as the quantity and impact.

Lord West – fascinated by AS' presentation. When a Minister put a lot of pressure on providers around paedophilia and got a lot of traction, same level of traction not there for women and girls, finds this worrying.

Big providers have trapdoors to decrypt messages. Useful for us to have good encryption, need to be aware of who can de-crypt. Need a mechanism for the good actors (who are they?) to get access.

AS – correct that there has not been traction nor the same level of thinking about this. Certainly not enough on child abuse. What are the harms, what is the likelihood to suffer online harms? Cannot impress enough the need to identify intersectionality and the impact on users of online platforms.

On facial recognition a lot of the equality arguments get lost. Black people are at a higher risk of mis-identification using facial recognition technology. Need to look at the whole effect.

KM – someone mentioned that a backdoor cannot be used if the encryption is used properly. Talking about encryption that no-one has access to e.g. FBI vs Apple[1]. In the West tends to be open better known. If good actors have access to the data, then the bad actors can as well. Complicated question.

Lord West – we are particularly bad about protecting data. Need to get better at this.

SF – subscribes to KM's view, what is the solution. If you weaken security, that is dangerous.

SR – solution is in metadata. Law enforcement analyses the metadata which cannot be hidden. Giving backdoors to encryption is a slippery slope. You can only have one back door in your encryption. If the UK Govt asks for access to the back door, then the US, then the French etc. so any legitimate government could ask for access. Not all legitimate governments are benign so how does FB make the rules.

Paul Simmonds – how do you make people accountable? You will always get situations where husbands find out which refuge their wife is in for example, you can call the police who will arrest the husband. Like banking, the platforms need to know their customers like Banks do. Social media IDs are self-asserted.

Rachel O'Connell – DCMS and the Home Office took the view that if platforms knew the age band of the user in a reliable way they would take action and build this in. Ran trials with the Football Association and proved the point so platforms can afford special protections for age related content. Used to work for BEBO as CISO and has done work around identity for BSI. An identity layer would be useful, like KYC in banks.

[1] F.B.I. Asks Apple to Help Unlock Two iPhones - The New York Times (nytimes.com)

Elizabeth Denham – 3rd party providers should provide this type of KYC, lower than that needed for a bank account for instance. Is it correct to encrypt communication between children? We treat car safety differently depending on the age group, air bags being the example.

Prof. Keith Mayes – back doors imply a weakness in cryptography, this is wrong. You should talk about an intercept service, this is the proper approach to this problem. We need a facility for de-crypting.

If you really want to tackle some of the problems that KM has mentioned, you need access to raw information in order to measure it. Has to be done in an authorized way.

Agree with the comment on identity. Platforms need a stronger link to user's real ID.

Use of covert devices – a lot of apps sneak in tracking. One way that you might do that with a GPS locator, when you give permissions, this should not be forever. The user should be given the option to renew the tracking facility. A lot of tracking goes on in cookies. You need access to raw data to do the analytics.

David Rennie – hope in this context that a different approach is taken. Should look on it as a societal good to prevent online harms. Needs to be looked at more holistically.

Conclusions –

AS – a lot of the conversation has been around encryption. This has become a selling point for many companies. At an early or commissioning phase during design, need to build in mitigation of potential harms.

SR – shining a spotlight on the question of these harms is very important. Avoid trading one problem for another. Needs wider education about these problems, help people to understand what they can do, give access to help to counter abuse (reclaim the streets).

KM – big issue for our society. SR's mention of meta data is a good point. As humans we are creatures of habit and even meta data can reveal an individual's identity. This is a problem for surveillance.
I think that there was a general consensus around the design of platforms, their architecture and services. Goes back to what KM mentioned about cookies. If we can nudge, not regulate, the companies themselves that would be a great step in the right direction. Education is also good. Look at the pandemic, lots of education but people do not necessarily follow the advice.

SF – thanked the speakers for their time and thoughts. The minutes will be circulated and a note sent to the relevant Minister.

Next meeting – 15th June DIT – expanding the UK's trade in cyber security