Take care with that social media duty of care

Should social media platforms be subject to a statutory duty of care, akin to occupiers' liability or health and safety, with the aim of protecting against online harms? In a series of blogposts and evidence to the House of Lords Communications Committee William Perrin and Professor Lorna Woods suggest that the answer should be yes. They say in their evidence:

"A common comparison is that social media services are "like a publisher". In our view the main analogy for social networks lies outside the digital realm. When considering harm reduction, social media networks should be seen as a public place – like an office, bar, or theme park. Hundreds of millions of people go to social networks owned by companies to do a vast range of different things. In our view, they should be protected from harm when they do so.

The law has proven very good at this type of protection in the physical realm. Workspaces, public spaces, even houses, in the UK owned or supplied by companies have to be safe for the people who use them. The law imposes a "duty of care" on the owners of those spaces. The company must take reasonable measures to prevent harm."

The aim of this post is to explore the comparability of offline duties of care, focusing on the duties of care owed by occupiers of physical public spaces to their visitors.
From the earliest days of the internet people have looked to offline analogies in the search for legal regimes suitable for the online world. Book and print distributors, with their intermediary role in disseminating information, were an obvious model for discussion forums and bulletin boards, the forerunners of today's social media platforms. The liability of distributors for the content of the materials they carried was limited. The EU Electronic Commerce Directive applied a broadly similar liability model to a wide range of online hosting activities including on social media platforms.
The principle of offline and online equivalence still holds sway: whilst no offline analogies are precise, as far as possible the same legal regime should apply to comparable online and offline activities.

A print distributor is a good analogy for a social media platform because they both involve dissemination of information. However, the analogy is not perfect. Distribution lacks the element of direct personal interaction between two principals who may come into conflict, a feature that is common to both social media and a physical public place. The relationship between a social media platform and its users has some parallels with that between the occupier of a physical space and its visitors.

A physical public place is not, however, a perfect analogy. Duties of care owed by physical occupiers relate to what is done, not said, on their premises. They concern personal injury and damage to property. Such safety-related duties of care are thus about those aspects of physical public spaces that are less like online platforms.
That is not to say that there is no overlap. Some harms that result from online interaction can be fairly described as safety-related. Grooming is an obvious example. However that is not the case for all kinds of harm. It may be tempting to label a broad spectrum of online behaviour as raising issues of online safety, as the government has tended to do in its Internet Safety Strategy Green Paper. However, that conceals rather than addresses the question of what constitutes a safety-related harm.

As a historical note, when a statutory duty of care for occupiers' liability was introduced in 1957 the objective was to abolish the fine distinctions that the common law had drawn between different kinds of visitor. The legislation did not expand the kinds of harm to which the duty applied. Those remained, as they do today, limited to safety-related harms: personal injury and damage to property.

Other closer kinds of relationship, such as employer and employee, may give rise to a duty of care in respect of broader kinds of harm. So under the Health and Safety Act 1974 an employer's duty in respect of employees is in relation to their health, safety and welfare, whereas its duty in respect of other persons is limited to their health and safety. The employer-employee relationship does not correspond to the occupier-visitor relationship that characterises the analogy between physical world public spaces and online platforms.

Non-safety related harms are generally addressed by subject-specific legislation which takes account of the nature of the wrongdoing and the harm in question.
To the extent that common law duties of care do apply to non-safety related harms, they arise out of relationships that are not analogous to a site and visitor. Thus if a person assumes responsibility to someone who relies on their incorrect statement, they may owe a duty of care in respect of financial loss suffered as a result. That is a duty owed by the maker of the statement to the person who relies upon it. There is no duty on the occupier of a physical space to prevent visitors to the site making incorrect statements to each other.

Many harms that may be encountered online (putting aside the question of whether some are properly described as harms at all) are of a different nature from the safety-related dangers in respect of which occupier-related duties of care are imposed in a physical public space.

We shall also see that unlike dangers commonly encountered in a physical place, such as tripping on a dangerous path, the kind of online harms that it is suggested should be within the ambit of a duty of care typically arise out of how users behave to each other rather than from interaction between a visitor and the occupier itself.

Duties of care arising out of occupation of a physical public place

The "operator" of a physical world place such as an office, bar, or theme park is subject to legal duties of care. In its capacity as occupier, by statute it automatically owes a duty of care to visitors in relation to the safety of the premises. It may also owe visitors a common law duty of care in some situations not covered by the statutory duty of care. In either case the duty of care relates to danger, in the sense of risk of personal injury or damage to property.

The Perrin/Woods evidence describes the principle of a duty of care:

> "The idea of a "duty of care" is straightforward in principle. A person (including companies) under a duty of care must take care in relation to a particular activity as it affects particular people or things. If that person does not take care and someone comes to harm as a result then there are legal consequences. ...
>
> In our view the generality and simplicity of a duty of care works well for the breadth, complexity and rapid development of social media services, where writing detailed rules in law is impossible. By taking a similar approach to

> corporate owned public spaces, workplaces, products etc in the physical world, harm can be reduced in social networks."

The general idea of a duty of care can be articulated relatively simply. However that does not mean that a duty of care always exists, or that any given duty of care is general in substance.

In many situations a duty of care will not exist. It may exist in relation to some kinds of harm but not others, in relation to some people but not others, or in relation to some kinds of conduct but not others.

Occupiers' liability is a duty of care defined by statute. As such the initial common law step of deciding whether a duty of care exists is removed. The statute lays down that a duty of care is owed to visitors in respect of dangers due to the state of the premises or to things done or omitted to be done on them.

"Things done or omitted to be done" on the premises refers to kinds of activities that relate to occupancy and create a risk of personal injury or damage to property – for instance allowing speedboats on a lake used by swimmers, or operating a car park. The statutory duty does not extend to every kind of activity that people engage in on the premises.

The content of the statutory duty is to take reasonable care to see that the visitor will be reasonably safe in using the premises for the purposes for which he is invited or permitted by the occupier to be there. For some kinds of danger the duty of care may not require the occupier to take any steps at all. For instance, there is no duty to warn of obvious risks.

As to the common law, the courts some time ago abandoned the search for a universal touchstone by which to determine whether a duty of care exists. When the courts extend categories of duty of care they do so incrementally, with close regard to situations in which duties of care already exist. They take into account proximity of relationship between the persons by whom and to whom the duty is said to be owed, foreseeability of harm and whether it is fair, just and reasonable to impose a duty of care.

That approach brings into play the scope and content of the obligation said to be imposed: a duty of care to do what, and in respect of what kinds of harm? In *Caparo v Dickman* Lord Bridge cautioned against discussing duties of care in abstract terms divorced from factual context:

> "It is never sufficient to ask simply whether A owes B a duty of care. It always necessary to determine the scope of the duty by reference to the kind of damage from which A must take care to save B harmless."

That is an especially pertinent consideration if the kinds of harm for which an online duty of care is advocated differ from those in respect of which offline duties of care exist. As with the statutory duty, common law duties of care arising from occupation of physical premises concern safety-related harms: personal injury and damage to property.

Outside the field of occupiers' liability, a particularly close relationship with the potential victim, for instance employer and employee or school and pupil, may give rise to a more extensive duty of care.

A duty of care may sometimes be owed because of a particular relationship between the defendant and the perpetrator (as opposed to the victim). That was the basis on which a Borstal school was held to owe a duty of care to a member of the public whose property was damaged by an escaped inmate.

Vicarious liability and non-delegable duties of care can in some circumstances render a person liable for someone else's breach of duty.

However, none of these situations corresponds to the relationship between occupiers of public spaces and their visitors.

A duty of care to prevent one visitor harming another

An occupier's duty of care may be described in broad terms as a duty to provide a reasonably safe environment for visitors. However that bears closer examination.

The paradigm case of a visitor tripping over a dangerous paving stone or injured when using a badly maintained theme park ride does not translate well into the online environment. The kind of duty of care that would be most relevant to a social media platform is different: a duty to take steps to prevent, or reduce the risk of, one site visitor harming another.

While that kind of duty is not unheard of in respect of physical public places, it has been applied in very specific circumstances: for instance a bar serving alcohol, a football club in respect of behaviour of rival fans or a golf club in respect of mishit balls. These related to specific activities that created the danger in question. The duties apply to safety properly so called - risk of personal injury inflicted by one visitor on another – but not to what visitors say to each other.

This limited kind of duty of care may be compared with the proposal in the Perrin/Woods evidence. It suggests that what is, in substance, a universal duty of care should apply to large social media platforms (over 1,000,000 users/members/viewers in the UK) in relation to:

> "a) Harmful threats – statement of an intention to cause pain, injury, damage or other hostile action such as intimidation. Psychological harassment, threats of a sexual nature, threats to kill, racial or religious threats known as hate crime. Hostility or prejudice based on a person's race, religion, sexual orientation, disability or transgender identity. We would extend the understanding of "hate" to include misogyny.
> b) Economic harm – financial misconduct, intellectual property abuse,
> c) Harms to national security – violent extremism, terrorism, state sponsored cyber warfare
> d) Emotional harm – preventing emotional harm suffered by users such that it does not build up to the criminal threshold of a recognised psychiatric injury. For instance through aggregated abuse of one person by many others in a way that would not happen in the physical world ([...] on emotional harm below a criminal threshold). This includes harm to vulnerable people – in respect of suicide, anorexia, mental illness etc.
> e) Harm to young people – bullying, aggression, hate, sexual harassment and communications, exposure to harmful or disturbing content, grooming, child abuse ([...])

f) Harms to justice and democracy – prevent intimidation of people taking part in the political process beyond robust debate, protecting the criminal and trial process ([...])"

These go far wider than the safety-related harms that underpin the duties of care to which the occupiers of physical world public spaces are subject.

Perrin and Woods have recognised this elsewhere, suggesting that the common law duty of care would be "insufficient" in "the majority of cases in relation to social media due, in part, to the jurisprudential approach to non-physical injury". However, this assumes the conclusion that an online duty of care ought to apply to broader kinds of harm. Whether a particular kind of harm is appropriate for a duty of care-based approach would be a significant question.

Offline duties of care applicable to the proprietors of physical world public spaces do not correspond to a universal duty of care to prevent broadly defined notions of harm resulting from the behaviour of visitors to each other.

It may be said that the kind of harm that is foreseeable on a social media platform is different from that which is foreseeable in a bar, a football ground or a theme park. On that basis it may be argued that a duty of care should apply in respect of a wider range of harms. However, that is an argument from difference, not similarity. The duties of care applicable to an occupier's liability to visitors in a physical world space, both statutory and common law, are limited to safety-related harms. That is a long standing and deliberate policy.

The purpose of a duty of care

The Perrin/Woods evidence describes the purpose of duties of care in terms that they internalise external costs and make companies invest in safety by taking reasonable measures to prevent harm. Harms represent "external costs generated by production of the social media providers' products".

However, articulating the purpose of duties of care does not provide an answer to how we should determine what should be regarded as harmful external costs in the first place, which kind of harms should and should not be the subject of a duty of care and the extent (if any) to which a duty of care should oblige an operator to take steps to prevent actions of third party users.

There is also an assumption that consequences of user actions are external costs generated by the platform's products, rather than costs generated by users themselves. That is something like equating a locomotive emitting sparks with what passengers say to each other in the carriages.

Offline duties of care do not attempt to internalise all external costs. Some might say that the offline regime should go further. However, an analogy with the offline duty of care regime has to start from what is, rather than from what is not.

Examples of physical world duties of care

It can be seen from the above that for the purpose of analogy the two most relevant aspects of duties of care in physical public spaces are: (1) the extent of any duty owed by

the occupier in respect of behaviour by visitors towards each other and (2) the types of harm in respect of which such a duty of care applies.

Duties owed to visitors in respect of behaviour to each other

One physical world example mentioned in the Perrin/Woods paper is the bar. The common law duty of care owed by a members' bar to its visitors was considered by the Court of Appeal in *Everett v Comojo*. This was a case of personal injury: a guest stabbing two other guests several times, leading to a claim that the owners of the club should have taken steps to prevent the perpetrator committing the assault. On the facts the club was held not to have breached any duty of care that it owed. The court held that it did owe a duty of care analogous to statutory occupiers' liability. The content of the duty of care was limited. The bar was under no obligation to search guests on entry for offensive weapons. There had been no prior indication that the guest was about to turn violent. While a waitress had become concerned, and went to talk to the manager, she could not have been criticised if she had done nothing.

The judge suggested that a club with a history of people bringing in offensive weapons might have a duty to search guests at the door. In a club with a history of outbreaks of violence the duty might be to have staff on hand to control the outbreak. Some clubs might have to have security personnel permanently present. In a club with no history the duty might only be to train staff to look out for trouble and to alert security personnel.

This variable duty of care existed in respect of personal injury in the specific situation where the serving of alcohol created a particular risk of loss of control and violence by patrons.

We can also consider the sports ground. In *Cunningham v Reading Football Club Ltd* the football club was found to have breached its statutory duty of care to a policeman who was injured when visiting fans broke pieces of concrete off the "appallingly dilapidated" terraces and used them as missiles. The club was found to have been well aware that the visiting crowd was very likely indeed to contain a violent element. Similar incidents involving lumps of concrete broken off from the terracing had occurred at a match played at the same ground less than four months earlier and no steps had been taken in the meantime to make that more difficult.

In a Scottish case a golf club was held liable for injuries suffered by a golfer struck by a golf ball played by a fellow golfer, on the basis of lack of warning signs in an area at risk from a mishit ball.

The Perrin/Woods evidence cites the example of a theme park. The occupier of a park owes a duty to its visitors to take reasonable care to provide reasonably safe premises – safe in the sense of danger of personal injury or damage to property. It owes no duty to check what visitors are saying to each other while strolling in the grounds.

It can be seen that what is required by a duty of care may vary with the factual circumstances. The Perrin/Woods evidence emphasises the flexibility of a duty of care according to the degree of risk, although it advocates putting that assessment in the hands of a regulator (that is another debate).

However, we should not lose sight of the fact that in the offline world the variable content of duties of care is contained within boundaries that determine whether a duty of

care exists at all and in respect of what kinds of harm.

The law does not impose a universally applicable duty of care to take steps to prevent or reduce any kind of foreseeable harm that visitors may cause to each other; certainly not when the harm is said to have been inflicted by words rather than by a knife, a flying lump of concrete or an errant golf ball.

Types of harm

That brings us to the kind of harm that an online duty of care might seek to prevent. A significant difference from offline physical spaces is that internet platforms are based on speech. That is why distribution of print information has served well as an analogy. Where activities like grooming, harassment and intimidation are concerned, it is true that the fact that words may be the means by which they are carried out is of no greater significance online than it is offline. Saying may cross the line into doing. And an online conversation can lead to a real world encounter or take place in the context of a real world relationship outside the platform.

Nevertheless, offensive words are not akin to a knife in the ribs or a lump of concrete. The objectively ascertainable personal injury caused by an assault bears no relation to a human evaluating and reacting to what people say and write.

Words and images may cause distress. It may be said that they can cause psychiatric harm. But even in the two-way scenario of one person injuring another, there is argument over the proper boundaries of recoverable psychiatric damage by those affected, directly or indirectly. Only in the case of intentional infliction of severe distress can pure psychiatric damage be recovered.

The difficulties are compounded in the three-way scenario: a duty of care on a platform to prevent or reduce the risk of one visitor using words that cause psychiatric damage or emotional harm to another visitor. Such a duty involves predicting the potential psychological effect of words on unknown persons. The obligation would be of a quite different kind from the duty on the occupier of a football ground to take care to repair dilapidated terracing, with a known risk of personal injury by fans prising up lumps of concrete and using them as missiles.

It might be countered that the platform would have only to consider whether the risk of psychological or emotional harm exceeded a threshold. But the lower the threshold, the greater the likelihood of collateral damage by suppression of legitimate speech. A regime intended to internalise a negative externality then propagates a different negative externality created by the duty of care of regime itself. This is an inevitable risk of extrapolating safety-related duties of care to speech-related harms.

Some of the difficulties in relation to psychiatric harm and freedom of speech are illustrated by the UK Supreme Court case of *Rhodes v OPO*. This claim was brought under the rule in *Wilkinson v Downton*, which by way of exception from the general rules of negligence permits recovery for deliberately inflicted severe distress resulting in psychiatric illness. The case was about whether the author of an autobiography should be prevented from publishing by an interlocutory injunction. The claim was that, if his child were to read it, the author would be intentionally causing distress to the child as a result of the blunt and graphic descriptions of the abuse that the author had himself suffered as a child. The Supreme Court allowed the publication to proceed.

The Court of Appeal had held that there could be no justification for the publication if it was likely to cause psychiatric harm to the child. The Supreme Court disagreed, commenting that:

> "that approach excluded consideration of the wider question of justification based on the legitimate interest of the defendant in telling his story to the world at large in the way in which he wishes to tell it, and the corresponding interest of the public in hearing his story. … "

It went on:

> "It is difficult to envisage any circumstances in which speech which is not deceptive, threatening or possibly abusive, could give rise to liability in tort for wilful infringement of another's right to personal safety. The right to report the truth is justification in itself. That is not to say that the right of disclosure is absolute … . But there is no general law prohibiting the publication of facts which will cause distress to another, even if that is the person's intention."

This passage aptly illustrates the caution that has to be exercised in applying physical world concepts of harm, injury and safety to communication and speech, even before considering the further step of imposing a duty of care on a platform to take steps to reduce the risk of their occurrence as between third parties, or the yet further step of appointing a regulator to superintend the platform's systems for doing so.

The Supreme Court went on to criticise the injunction granted by the Court of Appeal, which had permitted publication of the book only in a bowdlerised version. It emphasised the right of the author to communicate his experiences using brutal language:

> "His writing contains dark descriptions of emotional hell, self-hatred and rage, as can be seen in the extracts which we have set out. The reader gains an insight into his pain but also his resilience and achievements. To lighten the darkness would reduce its effect. The court has taken editorial control over the manner in which the appellant's story is expressed. A right to convey information to the public carries with it a right to choose the language in which it is expressed in order to convey the information most effectively."

Prior restraint

The Supreme Court in *Rhodes* emphasised not only the right of the author to tell the world about his experience, but the "corresponding public interest in others being able to listen to his life story in all its searing detail".

It may be thought that there is no issue with requiring platforms to remove content, so long as the person who posted it has access to a put back and appeal procedure.

That, however, addresses only one side of the freedom of speech coin. It does nothing to address the corresponding interest of others in being able to read it, a right which they will never be able to exercise if a platform has been required to prevent an item seeing the light of day and the originator then does nothing to challenge the decision.

We derive from the right of freedom of speech a set of principles that collide with the kind of actions that duties of care might require, such as monitoring and pre-emptive

removal of content. The precautionary principle may have a place in preventing harm such as pollution, but when applied to speech it translates directly into prior restraint. The presumption against prior restraint refers not just to pre-publication censorship, but the principle that speech should stay available to the public until the merits of a complaint have been adjudicated by a legally competent independent tribunal. The fact that we are dealing with the internet does not negate the value of procedural protections for speech.

Not every duty of care involves monitoring and removal of content. Not all use of words amounts to pure speech. Nevertheless, we are in dangerous territory when we seek to apply preventive non-specific duties of care to users' communications.

Duties of care and the Electronic Commerce Directive

Duties of care are relevant to the intermediary liability protections of the Electronic Commerce Directive. Article 15 prevents a general monitoring obligation being imposed on conduits, hosts or caches. However Recital 48 says:

"This Directive does not affect the possibility for Member States of requiring service providers, who host information provided by recipients of their service, to apply duties of care, which can reasonably be expected from them and which are specified by national law, in order to detect and prevent certain types of illegal activities."

This does not itself impose a duty of care on intermediaries. It simply leaves room for Member States to impose various kinds of duty of care so long as they do not contravene Article 15 or run counter to the liability protections in Article 12 to 14.

Article 15 again focuses attention on the question "A duty of care to do what?" A duty of care that required a user to have access to an emergency button would not breach Article 15. An obligation to screen user communications would do so.

Conclusion

This piece started by observing that no analogy is perfect. Although some overlap exists with the safety-related dangers (personal injury and damage to property) that form the subject matter of occupiers' liability to visitors and of corresponding common law duties of care, many online harms are of other kinds. Moreover, it is significant that the duty of care would consist in preventing behaviour of one site visitor to another.

The analogy with public physical places suggests that caution is required in postulating duties of care that differ markedly from those, both statutory and common law, that arise from the offline occupier-visitor relationship.