Social Media: How the Net is Closing in on Cyber Bullies

Jennifer Agate
Solicitor, Farrer & Co LLP

Jocelyn Ledward
Barrister, QEB Hollis Whiteman

LT Breach of confidence; Contempt of court; Criminal liability; Defamation; Harassment; Malicious communications; Social media; Threats

Summary

Online abuse can take many forms, whether it be the posting of a defamatory comment, a campaign of harassment or an isolated threat to cause harm. Whilst many perpetrators hide behind a perceived cloak of anonymity, recent action has demonstrated the potential consequences, ranging from named and shamed humiliation to jail sentences.

In this article we walk through the civil and criminal offences which can be committed (deliberately and unwittingly) via social media, examining the steps victims might take where they find themselves the target.

Cyberbullying

Online abuse hit the national headlines again in July this year when Caroline Criado-Perez, who had successfully campaigned for a woman's face to appear on UK bank notes, became the subject of a sustained campaign of abuse via social media site Twitter, including rape and death threats. In the same week other prominent women received similar abuse from so-called "trolls" via the social networking site. In August cyberbullying gained further prominence when 14-year-old Hannah Smith committed suicide after suffering bullying on anonymous social media site Ask.fm.

Following police investigations a number of arrests were made and Twitter has since responded with the introduction of a "report abuse" button.

These shocking examples of cyber abuse prompted widespread condemnation and discussion of attitudes and morality in today's society. However, it also became evident that despite the June publication of the final CPS Guidelines on prosecution cases involving communications sent via social media (the Guidelines), the boundaries of what you can and can't say and do online are far from clear. Similarly unclear is the dividing line between the responsibility of the user who posts on social media and the operator of the social media platform itself.

At first glance abuse sufficient to lead to suicide or rape or a bomb threat should surely give rise to some penalty. So where does the user risk a criminal prosecution or civil action for what they post online?

Threats

Threats of violence or damage to property can fall under a number of criminal regimes, including s.16 of the Offences Against the Person Act 1861 (threats to kill), s.4 of the Protection from Harassment Act 1997 (putting people in fear of violence), as well as s.1 of the Malicious Communications Act 1988 (threatening messages) and s.127 of the Communications Act 2003 (message of a menacing character). However, the Guidelines make it clear that to warrant criminal prosecution, the threats must be "credible".

In 2010 in the notorious "Twitter joke trial",[1] Paul Chambers was convicted of sending a "public electronic message that was grossly offensive or of an indecent, obscene or menacing character contrary to the Communications Act 2003". Chambers had tweeted about blowing Robin Hood airport "sky high" when the airport was forced to cancel flights in bad weather. The conviction attracted widespread public criticism and was ultimately successfully appealed. The Court of Appeal was particularly critical of the prosecution because it was clear that the threats were made in the context of a jokey exchange, not credible (and were never treated as such by the officials who were notified of them), and therefore lacked the necessary "menacing character".

"Credible" threats are more likely to be encountered where (unlike in *Chambers*) there is a real possibility that the threat will be taken seriously, such as when victim and perpetrator are known to each other, or in circumstances akin to anonymous "hoax calls" which have resulted in prosecution (e.g. bomb threats affecting town centres, public transport and airlines such as Matthew Davis, the BA steward who was jailed in August 2012 for writing a bomb threat on the back of an aircraft toilet door mid-flight, or by putative kidnappers during high profile missing person enquiries, as were made in relation to 12-year-old Tia Sharp, and ex-Eastenders actress Gemma McCluskie, both of whom had been murdered by relatives; in the latter case, Sam Dunne was sentenced to six months' imprisonment in April 2013).

The Guidelines state that as a general rule threats which are not credible should not be prosecuted unless they form part of a campaign of harassment targeting an individual within the meaning of the Protection from Harassment Act 1997 (see below).

[1] *Chambers v DPP* , [2012] EWHC 2157 (Admin) [2013] 1 Cr.App.R. 1

Abusive but not threatening comments

Communications considered to be "grossly offensive, indecent, obscene or false" could also fall under s.1 of the Malicious Communications Act 1988 and/or s.127 of the Communications Act 2003. However, to do so they will have to pass a high evidential threshold, because they involve restrictions to an individual's ECHR art.10 rights of freedom of expression, which must be necessary, proportionate and narrowly interpreted.

The observations of the Lord Chief Justice in *Chambers* as to what might be considered "grossly offensive, indecent or obscene" to an extent warranting criminal prosecution are adopted in the Guidelines: "satirical, or iconoclastic, or rude comment, the expression of unpopular or unfashionable opinion about serious or trivial matters, banter or humour, even if distasteful to some or painful to those subjected to it"[2] should not attract criminal sanction, even if "in bad taste, controversial or unpopular, and may cause offence to individuals or a specific community".[3]

The s.1 offence also requires proof that an indecent, grossly offensive, threatening or false communication was sent for the purpose of causing anxiety or distress. There is no statutory requirement to prove a specific intention or purpose under s.127, save in relation to "false" communications which must be for the purpose of causing annoyance, inconvenience or needless anxiety but see *Chambers* in relation to mens rea.

In October 2012 teenager Matthew Woods was sentenced to 12 weeks in a young offender institution after pleading guilty to sending by means of a public electronic communications network a message or other matter that was grossly offensive contrary to the 2003 Act.[4] Woods had posted explicit comments and jokes about missing (and ultimately murdered) child April Jones on his Facebook page. The sentence initiated an intense debate on whether the comments were simply repellent or criminal and worthy of prosecution. It is arguable (as many have) that under the new Guidelines, the case would not have satisfied the evidential test. The following month another Facebook user, Sam Busby, was charged under s.127 for making offensive remarks about the April Jones case. He received a six-week suspended jail sentence and was ordered to pay an £80 victim surcharge and keep to a 19.00–07.00 curfew for eight weeks.[5]

Where the abuse involves an element of race, religion or anything founded on disability, sexual orientation or transgender identity, the police and CPS are prepared to take the abuse particularly seriously. Part III of the Public Order Act 1986 may also be invoked (acts likely or intended to stir up racial or religious hatred, or hatred on the grounds of sexual orientation). In March 2012 university student Liam Stacey received a 56-day jail term for racially aggravated public disorder contrary to .31(1)(b) of the Public Order Act 1986 after tweeting "LOL" (laugh out loud) in response to the mid-match collapse of the footballer Fabrice Muamba and posting racist and offensive comments when other users criticised him for the original tweet. A subsequent appeal against the sentence was dismissed.[6]

Users of social media should also be aware that whilst the Guidelines do not encourage the indiscriminate use of public order legislation (such as Pt I of the Public Order Act 1986), as with many other statutes, they will be used in appropriate circumstances regardless of the fact that behaviour takes place online, and the user will not always be protected by the fact that he is acting from a private dwelling, rather than a public place. In 2012, Terry Balson was convicted of inciting others to riot after setting up Facebook group *For the Riot "Fuck the Feds"* which encouraged others to take part in the London riots in August 2011, although there was no evidence he had been present or physically participated in the riots.

In addition to the legal penalties, abusers can find themselves with longstanding reputational damage which will not only affect their social standing but their career prospects. In addition to the jail term (of which he served half), Stacey was banned from his university until the end of the academic year, although he was permitted to return as an external student to sit his examinations in the following year. In September this year a children's hospice nurse was suspended for six months by the Nursing and Midwifery Council for posting inappropriate messages on Facebook. The publicly accessible messages, which included a number of profanities and references to the hospice she worked for, were deemed likely to reflect badly on the nursing profession and therefore called her fitness to practise into question.

Photographs

The publication of private information without the consent of the subject could also give rise to civil action in the form of a privacy or breach of confidence suit. This would include the publication to third parties of private messages or photographs never intended to be shared with third parties, text messages and "sexts" and so-called "revenge porn" being common examples. Save where there is a public interest in publishing the private information, for example by exposing the hypocrisy of a married public role model engaging in an affair, or where the subject has themselves published the information, there will be limited defences.

[2] *Chambers* at [28]
[3] Guidelines at [39].
[4] *R v Woods* , Unreported November 1, 2012, Preston Crown Court.
[5] *R v Busby*.
[6] *R v Stacey* , Appeal No: A20120033.

In *Contostavlos v Mendahun*,[7] the singer Tulisa Contostavlos obtained an interim injunction against the dissemination of a leaked sex tape which had been made available online. Continuing an interim injunction, Mr Justice Tugendhat commented that

> "details of a person's sexual life have thus been recognised for very many years as high on the list of matters which may be protected …. It has also long been recognised that photographs are more intrusive than a verbal or written description. In the case of intrusive and intimate photographs of the kind in question in this case there is no real prospect of a defence of public domain".

The case subsequently settled.

In April an interim injunction was granted (continued in May) preventing the disclosure of photographs of a personal nature together with text messages sent by the claimant during the course of an adulterous affair with the second defendant. The photographs and text messages, described by the judge as "of a sexual nature, but could not be described as pornographic", had come into the control of the first defendant (with whom the second defendant had been in a relationship) in undisclosed circumstances. The court granted an order to protect the claimant's right to confidentiality and privacy, as well as to protect her from harassment.[8]

Where indecent images of children images are shared, as in the case of widely distributed photographs of an underage girl (in Ireland) pictured in a sexual act at a music concert in July, criminal offences may also be committed. There has yet to be a prosecution of a juvenile for taking or online sharing an indecent "selfie" and this may be considered highly unlikely (such an individual is far more likely to be regarded as a victim). CPS guidance on sexual offences generally encourages caution before commencing criminal proceedings against juveniles, and discourages criminal prosecution of underage genuinely consensual behaviour, but those who encourage the taking of such images or engage in wider sharing of them (particularly if the images are taken in circumstances which involve grooming, bullying, abuse or exploitation) should expect to be treated rather differently (even if juveniles themselves).

Contempt and breach of court order

Another area in which users can unwittingly find themselves in trouble is by publishing information online in breach of a court order or statutory prohibition.

A user will be in contempt of court if he/she publishes information in breach of a court order, most often in the circumstances of an injunction. In February the Attorney General prosecuted two social media users, Dean Liddle and Neil Harkins, in the first contempt proceedings concerning social media. Both Harkins and Liddle had published (on Twitter and Facebook respectively) photos purporting to represent the two killers of Jamie Bulger in breach of a longstanding worldwide injunction. After pleading guilty to contempt of court, both received suspended sentences of nine months.

Section 1 of the Sexual Offences (Amendment) Act 1992 gives the victims and alleged victims of rape and other sexual offences lifelong anonymity. In November 2012 nine individuals were fined for contravening this provision by publishing material likely to lead members of the public to identify the complainant in the Ched Evans rape case. They had also posted messages of abuse. From over 6,000 postings 21 arrests were made, with ten prosecutions. All received the maximum penalty and were required to pay £624 in compensation to the victim. The victim had to be given a new identity and was relocated.

Deliberations in the jury box are strictly confidential and must not be disclosed to anybody other than a fellow member of the jury. A juror discussing a case on social media therefore commits contempt of court. In July, the Attorney General successfully prosecuted two jurors Kasim Davey and Joseph Beard for contempt of court, both having been found guilty of misconducting themselves whilst serving as jurors in separate proceedings in the Crown Court. Mr Davey, who had been appointed as a juror in the trial of a sex offender, had posted a Facebook message to around 400 Facebook "friends" stating: "Woooow I wasn't expecting to be in a jury Deciding a paedophile's fate, I've always wanted to F**k up a paedophile & now I'm within the law!"

The court found that he had committed an act calculated to interfere with the proper administration of justice and which he intended would do so and sentenced him to two months, at least half of which was to be served. Mr Beard, whose case was heard at the same time, was also sentenced to two months for conducting internet research in breach of the guidelines given to jurors.

Comments made on social media might also fall foul of s.51 of the Criminal Justice and Public Order Act 1994 and ss.39 to 41 of the Criminal Justice and Public Order Act 2001, which create widely-drawn offences of intimidation and taking revenge against witnesses (actual or potential) in all proceedings before the Court of Appeal, High Court, Crown, County and Magistrates' Courts, jurors, and those who are assisting in criminal investigations. The fact that a comment is not made directly to the victim is immaterial. Users should bear in mind that *any act* which is intimidating and intended to be so would result in the commission of an intimidation offence; and whilst the revenge offence requires proof of an act or threat of harm, it may take place up to a year after the conclusion of proceedings, and the "harm" might be financial as well as physical.

[7] [2012] EWHC 850 (QB).
[8] *ABK v KDT* [2013] EWHC 1192 (QB).

Defamation

Moving away from those postings which do not cross the criminal threshold, are a raft of publications which can lead, sometimes unwittingly, to civil proceedings.

Where a user posts a statement of fact concerning a third party which lowers the subject's reputation in the minds of right thinking individuals, they risk a defamation suit. Sally Bercow discovered this when she tweeted her seven word question about Lord McAlpine, for which she ultimately paid damages and costs estimated at up to £3,000 per character. For further discussion see McAlpine, the Attorney General and the Defamation Act - Social Media Accountability in 2013 [2003] Ent. L.R. Issue 7 pp.233–235).

To bring proceedings, a claimant must first be prepared to take the risk of exposing the allegations to a wider audience by bringing them into open court and therefore qualified privilege protection for a potential publication. Once that decision is faced, under s.1 of the new Defamation Act 2013 they will need to show that the allegations caused them "serious harm", which in the case of companies must include evidence of financial loss. To how wide (and influential) a readership were the allegations published? Would they have been taken seriously? Are the words simply mere vulgar abuse?

Harassment (civil and criminal)

Where a user enters into a course of conduct involving any of the above, they could also face civil or criminal harassment proceedings.

The Guidelines encourage the use of the Protection from Harassment Act 1997 legislation where a particular individual is being targeted on social media, either by being put in fear of violence (the more serious offences under ss.4 and 4A), or by conduct that amounts to harassment, particularly through "stalking" behaviour under s.2A. Examples of "stalking" given in section 2A(3)(b) include contacting or attempted to contact a person by any means, and publishing any statement relating to a person or purporting to originate from them. Although a "course of conduct" must be proved, this may be shown by conduct "on at least two occasions" (in reality, a subjective assessment will be made of all the circumstances, and it is usual for conduct to be more extensive before it can be said that a reasonable person would be caused the requisite fear or distress).

In addition to the criminal offence, the Protection from Harassment Act 1997 also creates a civil statutory tort of harassment, allowing the victim to obtain a civil injunction and claim damages. In civil proceedings the standard of proof is on a balance of probabilities (lower than the higher criminal standard of beyond reasonable doubt).

Responsibility for publication

Users may also be unaware that by simply retweeting (or otherwise repeating) a statement originating from a third party, they face equal liability.

Where a defamatory statement has been made online, the operator of the website is able to rely on the defences under s.1 of the Defamation Act 1996 and Reg.19 of the E-Commerce (EC Directive) Regulations where they can show they were not on notice of the defamatory nature of the statements. Once properly notified they can become liable if they do not remove the material. *Tamiz v Google Inc*[9] confirmed that unless the content complained of is removed expeditiously by the website operator, it will be liable for the defamatory allegations.

Under s.5 of the Defamation Act 2013 (at the time of writing not yet in force), they will also have a further defence where they can show that on notification they followed a specified procedure by which the user is given the opportunity to stand behind their comments.

Some criminal statutes contain specific protection for providers,[10] but many do not. The Malicious Communications Act 1988 specifically defines the offending "sender" so as to include a "transmitter", but the practical reality is that if providers continue to co-operate with criminal investigations and with removal of content which may meet the criminal threshold, criminal prosecution of ISPs are unlikely to follow.

There is also a growing reputational pressure on ISPs through advertising (as Ask.fm found when advertisers pulled their adverts after the reports of Hannah Smith's suicide) and commercial pressure to co-operate with both police and civil claimants to overcome privacy and jurisdictional issues.

Threshold for criminal prosecution

For a criminal prosecution, a case must satisfy the test set out in the Code for Crown Prosecutors. First, the requirement of evidential sufficiency, and secondly the consideration of the public interest.

To pass the evidential stage, a prosecutor must be satisfied that there is sufficient evidence to provide a realistic prospect of conviction, i.e. that an objective, impartial and reasonable jury properly directed is more likely than not to convict. The Guidelines are clear, a case that does not pass this evidential stage must not proceed, no matter how serious or sensitive it may be.

Once the evidential threshold is passed, the prosecutor must consider whether a prosecution is required in the public interest.

The Guidelines make clear that those cases which amount to credible threats, breaches of court orders or specific targeting of an individual are likely to satisfy the public interest test and will be prosecuted robustly. But in relation to cases falling into the fourth defined

[9] *Tamiz v Google Inc* [2013] EWCA Civ 68; [2013] E.M.L.R. 14.
[10] SI 2010/894 is one example which protects those who are mere conduits, caching or hosting in relation to comments which amount to inciting hatred against persons on religious grounds or grounds of sexual orientation.

category, where communications are merely offensive, indecent, obscene or even false, many will not meet the public interest test. It is here where there is some overlap with the high evidential test for what is to be considered "grossly offensive".

In relation to this fourth category (but not the first three), whilst there is no exhaustive list of "public interest" considerations, genuine remorse, swift and effective action to remove offensive material, material that was not intended for a wide audience, or material which does not "go beyond what would conceivably be tolerable or acceptable in an open and diverse society which upholds and respects freedom of expression" will all be taken into account. Age and maturity are also relevant, and prosecution of those under 18 are rarely likely to be in the public interest.

Where will the line be drawn in practice? Each case must be considered on its own individual facts and merits. Although this was before the Guidelines had been published, the footballer Daniel Thomas did not face criminal charges for a homophobic Tweet about Olympic diver Tom Daley. Reasons given by the Director of Public Prosecutions in September 2012 were that it was a one-off message, not intended for wide circulation, swiftly removed and for which Thomas had expressed remorse. The case would undoubtedly have been decided in the same way if the Guidelines had applied. However, Thomas did not escape sanction entirely: he was suspended by the Football Association for one match and fined £500 later that year.

Some suggest that Liam Stacey would not have been prosecuted for his racist comments under the new Guidelines, particularly as he expressed immediate remorse and the trial judge found he had acted when drunk (although this is not usually treated as a mitigating feature by the criminal courts). It remains to be seen whether criminal charges will follow the arrests in relation to the Caroline Criado-Perez Twitter abuse. This is more likely to meet the criminal threshold because she has been targeted specifically and repeatedly for a prolonged period of time. If the predicted reduction in prosecutions which fall only into the "grossly offensive etc." category comes to fruition, we may not be able to gauge how high that evidential threshold is for some time.

Practical steps for victims

Although as above, there is the option of pursuing the website operator for defamatory comments once they are put on notice, to be held accountable, the individuals responsible for online abuse must first be identified, not always an easy task. In some cases the users will post under their own names, perhaps believing themselves to be safe from being held to account. In many cases, particularly where the user is outside the United Kingdom and jurisdictional issues arise, they are sadly often right.

In other more sophisticated campaigns of abuse the users will often attempt to protect their identities by using anonymous user names and accounts. A determined claimant can often obtain details of the user by pursuing a Norwich Pharmacal application in the courts, seeking a court order by which the website operator will be required to disclose the registration and log on details held for the relevant poster. However, these can be faked and IP addresses screened by the use of proxies, meaning that the costs of such application (generally anything from £5,000 upwards) may be expended in vain or which may require further applications.

To bring civil proceedings, a claimant must not only be prepared to face the cost of litigation, but must also be prepared to take the risk of exposing the allegations to a wider audience by bringing them into open court and therefore the possibility of reporting in the mainstream press.

For those who think they are or may be the victims of criminal conduct, the first step must always be to report matters to the police, even if they have no expectation that criminal proceedings will actually follow. The police have much wider powers and obligations under mutual legal assistance treaties to draw upon in order to identify perpetrators, and may be able to act more quickly. At the same time, all instances of contact (by whatever means) or postings about victims should be logged when seen—it may be a cumulative course of conduct which ultimately justifies arrest and charge, and it is notoriously difficult and sometimes impossible to access historical material from ephemeral social media such as Twitter, repeatedly edited blogs and even Facebook accounts. Do not shut down a social media account before consulting with the police, as vital evidence may be lost in the process. If unsatisfied with police response to a complaint, there is always the (costly) option of launching a private prosecution against a known perpetrator, but always bear in mind the powers of the DPP to take over any such prosecution and discontinue it if it is deemed not to be in the public interest.

The underlying issue

In the face of increasingly shocking headlines, an underlying issue exists. Rather than relying on legal remedies when abuse does take place, how do we halt this worrying trend towards cyberbullying and similar abuse? The online environment can give a slanted perception of reality, with the perceived anonymity and distance from the victim perhaps dulling the emotional reactions of the abusers. The safety of children online depends partly on educating them about what is safe to impart about themselves online (would you share the same information with a stranger offline) and partly by making internet users aware that what they do online will attract sanction in the same way as if they had done

it in another sphere (playground, public street, newspaper) and maybe more so given the greater permanency of electronically created material.