**INTIMIDATION OF PARLIAMENTARY CANDIDATES REVIEW**

**1:1 MEETINGS**

**NOTE OF MEETING WITH GOOGLE**

**Date: 2 November 2017**

**Time: 14:00**

**Location: Google’s office, Central St Giles, London**

**Present:**

* + Katie O’Donovan (KOD), UK Public Policy Manager, Google
	+ Lesley Bainsfair, (LB), Secretariat, *Committee on Standards in Public Life*
	+ Dee Goddard (DG), Secretariat, *Committee on Standards in Public Life*
	+ Stuart Ramsay (SR), Secretariat, *Committee on Standards in Public Life*
	+ By video conference; in New York, Yasmin Green and Lucy Vasserman,Jigsaw
	+ By video conference, in Paris, David Skelton (DS), UK Public Policy Manager, Google

*The meeting was conducted on the basis that we would make an audio recording of this meeting and will use this for quotes in our final publication. We have since requested publication of the full transcript. on our website as we have with our meetings with other social media companies and political parties.*

LB: I am Lesley Bainsfair, the Secretary to the Committee on Standards in Public Life, and Dee Goddard and Stuart Ramsay are Senior Policy Advisors to the Committee. Apologies from Lord Bew and Jane Ramsey who are unable to join the meeting. We thought it best to go ahead with the meeting as it has been difficult to set up. Thank you for David’s evidence last week and for following up with further information.

We were asked by the Prime Minister in July to review the intimidation of Parliamentary Candidates, looking at their particular experience in the UK June general election this year, and the wider implications for those in public life. Obviously, we want to consider the role of social media in that.

KOD: The technical people will be ready from 2.15 and Dave is available now. Let’s dial Dave in and the Perspectives team can join very shortly.

LB: Let’s start with Jigsaw. In July this year, the Prime Minister asked the Committee to undertake a review into the intimidation of Parliamentary candidates, especially the recent UK General Election. We intend to produce the report before Christmas. We’re very grateful to David who we saw last week, but we are keen to hear more about your work. If you would like to say a few words, and then Dee can come in with some questions?

YG: I’m Yasmin Green.

LV: I’m Lucy and I’m a software engineer at Jigsaw.

YG: Can I give you 60 seconds on Jigsaw[[1]](#footnote-1)? We play a special role in the Alphabet world. We’re relieved of having any commercial goals, or anything near-term on responding to the company. Six years ago Eric Schmidt, the then-CEO, wanted to have a team thinking about the next generation of people coming online and the issues they will bring with them – there was thinking about [...], Iraq, Syria and we ended up thinking about technology. The relationship between geopolitics and technology is only growing and we need to have a longer time horizon. Our role is looking at the ability of technology to protect people from these security threats. Our priorities are defending against cyber attacks including denial of service attacks, repressive censorship, online hate, and online harassment and radicalisation. We look at all these things which are not part of any part of what any technological platform’s vision for what online should look like.

LB: Sorry, Yasmin, I want to make clear as we said to Katie that we would make an audio recording of this meeting, as we have done with our other meetings.

YG: We spend a lot of time thinking about the tension between promoting free speech – making information freely available to people – and on the other hand protecting people from harmful speech.

You’ll see across tech platforms, what we are trying to grapple with, is that we are seeing more cases of malicious use of the internet. We need to protect users, even from a commercial perspective, we need to make people feel safe online. We would like to talk to you more broadly, because politicians and journalists are two of the user archetypes that really inform our work. We were thinking initially of human rights activists that it was obvious that oppressive regimes would target, but it became obvious journalists and politicians are at the frontline of these attacks and can inform us the state of the art about how the internet is being exploited to harass people. We do have questions for you about what you are seeing to help inform our work. Thank you for the function that the Committee is playing.

[Redacted]

Because there is so much content being uploaded, we can’t do as much as people want us to do. It became an unsatisfactory answer and increasingly unsatisfactory as more content was uploaded. Several years ago we made a bet and this is really early stage - this is a really ambitious technology - we invested a lot in Conversation AI. The idea was that we can use the ability of machines to understand context, to understand language. Maybe we can we actually teach algorithms to understand context. Maybe we can help these communities online who are struggling to have civil conversations, maybe we can give them the tools to do so. Until now platforms have typically have one set of terms of service, a universal truth, there is one definition of bad speech. But as it universal it errs on the side of being conservative, of promoting free speech, but doesn’t prohibit all intimidation. We think that technology should be able to allow different communities different types of conversations. I might want to dance in a moshpit at a rock concert, but I’m going to behave differently at a wedding. There are many versions of ‘me’, and the physical world is an analogy for me being different in different places with different rules. Can we try to let different communities in different parts of the internet have different notions of civility or what is acceptable. Lucy is on the tech team to is handling the tech side. We can talk about what that looks like on the product too but Lucy can talk about the technology.

LV: We have many, many, hundreds of thousands of comments shared with us by places like the New York Times, Wikipedia, many news sites. Some like NYT already do moderation - this is what we want, this is what we don’t want. Most of the time we have them relabelled by a crowdsourced army of humans who we hire, and we ask them ‘is this rude or disrespectful, would this make you want to leave the conversation?’ We get a lot of different people’s opinion on that.

We put that into a machine learning model that’s going to make predictions. You look at a piece of text and it’s going to give you a number that represents how likely it is that someone will perceive this as rude or disrespectful or make you want to leave the conversation. That machine learning model, that’s the first model, but we’re working on building a lot of models that do different versions of the same problem, like is this inflammatory, is this obscene, is this hate, things like that. That’s our base, that’s the base technology we have and we have that in a public API, which means that anybody can just send us a piece of text and we will send them back this number. And that’s open to be used for any product, as people see fit. And we have lots of opinions and advice on what is the right way to use it, and we are encouraging more partnering with certain people to build tools around this technology. For example, for the New York Times we have a whole moderation system which sorts the comments which people are typing, which they’re building around these sorts of models which helps moderators make decisions. Another case is called Authorship – some websites, as they are typing it that comment, it goes to our API at that moment, and that can give the user a heads up, ‘you might want to think twice before you send this comment’.

DG: That’s really interesting. Is anyone using this userside, that you just mentioned? The viewer author experience?

LV: There’s this platform called Vuukle – they do the comments for several thousand websites and they use it.

DG: We tried it out ourselves, and it gives you a traffic light system, of the perceived toxicity. We’ve seen you have different partners investing in the project. Does Google intend to use the API in their comments sections perhaps on YouTube? How much is it embedded into the Google side of things?

LV: The user team has been working on similar issues for a long time, so we collaborate with them in various ways. I don’t think they have any plans to add Authorship experience, just like that, but we are in touch with them.

DG: Could you use this as a way to supplement moderation, for example, to escalate complaints. To what extent has that been embedded? Is that something that is being worked on?

KOD: Yasmin and Lucy work in the Jigsaw part of Google. They will create the products in tandem and in relation with Google’s different products, but there’s no immediate integration with other products in our family.

LV: All of the other teams get to choose what they include. For YouTube, they are building their own tools on this but we are heavily collaborating with them.

YG: There’s this question of integration, how does this technology appear in products, versus trying to improve the technology. You’ve said you’ve tried it yourself so you will know it’s not perfect – so we’re doing a lot of work to improve the accuracy - we don’t want false positives or negatives but we are collaborating to improve the technology. Also the technology itself is based on TensorFlow which is a machine learning technology across Google. In terms of product integrations we’re thinking more outside of Google – that is just our mandate.

DG: That is helpful clarification. What we are trying to do is build a picture of how things are, particularly around elections, and how things could be, and understanding how you are working with Google is quite helpful for us in figuring out how things are playing out on social media. Could you give an outline of how it’s being implemented at the moment? Or how you intend that the API might be implementing?

LV: It’s already implemented – you already playing with it so that’s a public API. Are you familiar with machine learning? It’s a machine learning algorithm, you give a lot of comments with the correct answer that you want the model to learn. It’s a very large statistics system. And it says ‘these three words mean this, or these ten words in this order, that means this’, and so it learns associations and can then do predictions for you.

YG: The API is what we are making available. I said we’re not responsible for making money for Google and we’re not confined to Google’s platforms. I don’t think our effort with conversation AI can be successful unless we work hand in hand with lots of other platforms across the internet. We want to model, because speech looks different – as you will know intimidation can look very different across platform. For example, different political leanings and the way that speech looks different at different places across the internet. The API is making available the scoring system for people to try, and for developers to make something. There are thousands of developers who are using or applied to use the API in something they’re developing

[Redacted].

And we’re working with partners, so New York Times are using with all their moderators. When we started working with them, I’d need to check this but I think 10% of their homepage had comments enabled because it was so laborious to review the comments and have a civil conversation – with the help of Perspective, NYT have been able to double the number of articles they allow comments on. This is the only example we know of publishers wanting to enable speech on more on their content, rather than less. The story of the last couple of years was ‘I don’t have the resources, I don’t have the expertise to make sure that they have a civil conversation’, so whether it is Reuters and others, they would just turn comments off. It’s a tremendous challenge for free speech and how we solve it will change how the internet will look.

DG: Thank you. I think most of the other questions we have are about policy. So is there anything further about the future direction you see automation going in to combat hate speech online?

LB: You talked about its different applications, and we’re obviously interested in intimidatory or abusive speech. Do you view that as a separate project or do you see it falling under the same area?

LV: Yes. In the very first systems were building we’re just looking for toxicity, we call it. We believe that harassment and intimidation most of the time falls under that bucket, and most people say when they saw that, ‘this is rude and disrespectful and would make us want to leave the conversation’, and that’s our definition of toxic. Since then we have being doing more personalisations so that people can customise what they want for their platform. We are trying to build more specific models, focussing on breaking down toxicity into its components e.g. obscene language, insults, threats and identity-based hate and those kind of things, and for other people it might make sense to allow obscene words whilst not allowing aggressive intimidation. That’s a little bit more new and that is where we are going in the future – separating out different aspects of what is offensive online.

LB: Do you have any sense of how long this development takes, what is the timeline?

LV: We already have a very preliminary model of…are you asking specifically about hate speech?

LB: Yes.

LV: We have a very preliminary model that is in alpha right now, I think it is called identity-based hate. It might be so new that it’s not yet in our documentation but I think you can access it. So, the way that we work when we start working on this is that we put out what we have but they can have something very quickly…whether it’s something we feel comfortable using. We are not anywhere close to that yet – we are just beginning still and can’t say when we’ll get there. In the next, maybe, six months or so, that’s something we’re trying to prioritise for the next couple of months so we’ll have an idea of how it’s going.

LB: Is it a large team working on the hate speech issue or a small number of researchers and engineers?

 [Redacted]

KOD: We have a spam and abuse team which sits across Google with an engineering team that looks at issues from hate speech through to spam, so you have the engineering contingent there, and each product and platform will have its own engineers and others who are working on these issues, so that will range from lawyers and policy people to trust and safety practitioners and more engineers there, so different teams across Google. Jigsaw is an adjunct to our own products and platforms in that sense.[[2]](#footnote-2)

YG: Thinking about our model, this isn’t Google Alphabet trying to solve problems for everyone else but working together to do it because we have to work collaboratively, on the data side and also the product quality side. That’s why it would be amazing if we could have continuing relationship with your Committee or a more appropriate group in the UK for us to work with. Our model is by definition not real, not realistic, it’s a simplification, so there will be some level of inaccuracy or bias. That’s OK, but for a group that has a mandate to build more inclusive conversations we can’t be systematically missing the abuse that politicians or politicians of particular descriptions get.

[Redacted]

We see this as we can provide technological powerhouse but from a data/product quality and policy perspective we need input from lots of different groups.

LB: I see that’s the case. Do you work with other technology companies?

YG: No we don’t, I don’t know what they’re doing. I don’t know about anyone else that has the backing of an enormous technology company that has the lens of the whole internet.

LB: Thank you.

YG: Thank you very much for your work.

LB: That was really helpful, especially on the technological side. We had a few more questions following the conversation with David last week.

DG: Is Google in a position to bolster its monitoring of threats or abuse including in the run up to election campaigns? Do you think there is any additional responsibility in the run up to elections to protect the processes of Parliamentary democracy?

KOD: When you say Google, are you talking about, say, YouTube?

DG: We were talking with David about this as well. We are talking about YouTube - there’s a lot of political activity that goes on on YouTube - but also a responsibility for information online, in terms of search, in terms of making sure hateful stuff doesn’t come up so high in rankings, so more across the piece.

KOD: On search, we want to deliver the most meaningful and accurate result that a user is looking for. So during election times, that is obviously key moments with people understanding who are they going to vote for and what are the issues matter to them, and for most people that’s the sole time of doing that. So we make sure the election service alongside search is accurate and robust and provides balanced information and that people can trust what they get and we are in no way a directive platform. The way that search functions is that we deliver what is available on the web and we use an algorithm to deliver the most accurate information that people are seeking. Trustworthiness is really important. So if you were to search for ‘what does the budget mean for me’, post-budget, we would use measures within the algorithm to identify what are the most trustworthy sources talking about the budget, because people want the most accurate information, you know, ‘will I be worse off’. So there are functions the algorithms will detect to make sure that these are the trustworthy sites. On YouTube we see more activity around elections on politics, and we generally think that’s a good thing something that can be encouraged. That’s people participating, creating content, and reviewing and distilling content on there that might be produced.

In terms of those creating content, I was talking with David about this earlier in the week, the last election we had here was a snap election, certainly that we didn’t know about and a lot of candidates didn’t know about too. I think they were using new areas of technology, and I think YouTube is less intuitive than other platforms for politicians to use . We have a role in re-educating politicians first about what YouTube is and how they can create content but also how they can help engage with other content that might be on there. Some of that might about bespoke advice for MPs. In other countries, we’ve been able to provide really in depth support and information to MPs and candidates so they can use the platform. For any content creator once you have uploaded something on to YouTube it’s completely up to you to decide how you distribute that, might chose to keep it private to those that you share the link with like friends or family, or supporters; or public for everyone. Then you have the opportunities to control the comments and messages that appear underneath that. So can have no comments, or you can pre-vet comments, whether every single comment or particular keywords for review.

[Redacted]

So there is nuance and control there but not everyone knows about that. And that’s something we would definitely want more people to use. If you want you can pre-vet every single comment. For example, you can have your own decision of robust debate on a bypass issue but you can filter out personal abuse that you don’t want there. The application of the policies and the community guidelines that we have is really really key. We have the tools there for people to do it personally, but also have the community guidelines so that if people do want to flag comments which are hateful or incite hate or are abusive, then we will review those. Not sure if this will come out in your report, or if you can share information now, but the way that YouTube is used in this space compared to other platforms. We’re part of MOPAC and they have a hate crime hub with the Met police and wider partners,, which we’ve recently joined but we haven’t been long-term members. They have a project where the Met Police on a London-wide basis, if you phone the Met Police and you say I’ve been subjected to online hate. If there’s a criminal bar, then obviously the police pursue that but if a non-criminal but still substantive issue then the police contact Stop Hate UK who will asses and then can direct that to the relevant social media. I haven’t got official figures to share with you but the latest data I have is very few or no instances on YouTube, which is by no means to say that we don’t have toxic...we have everything from abusive, to toxic, to robust debate. But we aim to do with YouTube is to give people a platform where they can create content and reach people to engage with content with other means. It’s a very general use platform from that perspective. And we would work very hard to preserve that space, for politicians and others, to upload or share things, hyper-local videos that wouldn't get on to national news, or political dialogue. We want it to be a place where there is robust debate but not hate speech.

DS: Can I jump in to go back to the earlier bit about search during elections. We publish data which makes it clear that Google searches spikes during campaigns. On Google News, we think it’s particularly important at all times, but especially during elections that there is a diverse of opinion and news sources. But something we’ve been introducing over the past 6-12 months is that including fact checking of news – if you search then results will come up but also the likes of Snopes and Full Fact will appear in Google News. If you search for disputed issues there will be a fact check. We think it’s really important that during election campaign that people have tools to engage in those campaigns.

LB: You were talking about MOPAC and cases that fall short of illegality but that are inappropriate and that are flagged to you. How do you respond to that and what is the process?

KOD: We have flagging systems for both content and comments. They will be triaged, particularly for content, depending on what you are flagging for. So you have everything from copyright violation which is treated very differently to a privacy violation or hate speech. The same happens for a comment. It goes through a triage system for a review if the automated triage system isn’t able to decide the right action. We have teams that sit around the world – the nearest one to the UK is in Ireland – and they will review that and they will take an action. Some comments that flagged will stay up, because they don’t break the community guidelines. For example, people will flag comments that they disagree with, like ‘Vote Labour’ or ‘Vote Conservative’, we won’t remove that. But when it reaches the bar that crosses our community guidelines which does prohibit hate speech, then we will remove the comment.

LB: How long does that take?

KOD: We don’t regularly or routinely publish figures for that, but it happens on quite an efficient process. We aim to do it as quickly as possible.

DG: Is there a different process for content which is perceived to be intimidatory or which breaks the rules? We have heard examples of MPs who have had been subject to content uploaded which is shared widely around their local area but then goes around and would affect the election. Once it was identified it wasn’t taken down before the election. There are a couple of weeks where it is flagged but hasn’t been taken down. Election campaigns are distinct in intensity – that’s true about the robustness of debate at that - but also how time really matters.

KOD: We have a programme of trusted flaggers. They tend to be subject matter experts, for example, a support charity, on content relating to suicide or a child protection charity on content relating to children. They tend to be, but don’t exclusively have to be, charities or experts. I don’t think we have any that are political parties and I think we would have to think carefully about the process and whether it was the individual political party that was the relevant organisation to be the trusted flagger or perhaps an Electoral Commission or a more neutral space. But we would be very happy to work with trusted flaggers of the appropriate kind. What the trusted flagger can do, is that they can become an expert in the content that is not allowed on our platform, and they can flag that. We can understand that they are an individual or an organisation who has a very high understanding of our processes, and that they know when to alert us, and that can help us get an expedited review and also help to feedback to them about the processes. That would be something well worth exploring. It would be about who the organisation was. It might be that we do, in other countries, work with political parties but it might be that a more neutral but still accessible organisation might...

DG: With trusted flaggers, the content it isn’t taken down right away? Only expedited through the process?

KOD: Yes, even internally we will have content where we look at it and we have to have a really good internal analysis. Sometimes issues are just so nuanced that actually giving that responsibility to a third party is very difficult. We’ve found the most effective way to do that is to keep the responsibility for determinations about our community guidelines in-house but absolutely have that expertise able to flag that content.

LB: One of the MPs we spoke to, has understood that viewpoint which has been expressed by social media companies, but actually takes the opposite view and said ‘this is very clearly hate crime, why isn’t the social media company’s response to take it down immediately and then review it. Why do you decide to leave up until you decide it has crossed a particular line?’

KOD: I would hesitate to give a conclusive answer to that. A lot of our policies are formed on a broad global basis. We then do localised policies – so in the UK some content is illegal that in other countries isn’t. And there is content in Germany that is illegal there but isn’t here. We respect local law[[3]](#footnote-3). It would be good to think more about this question. In other countries we have seen periods of political unrest or political change – where at the same time, associated with it, we have seen enormous spikes of the flagging of content on YouTube by government organisations. It seems slightly ridiculous sat here in the UK where we have very strong and robust Parliamentary processes and laws that govern what we do, but there would be a hesitation in handing that responsibility to an organisation where over time the role and influence of that organisation might shift. And the other thing we’ve noticed is that when we have a policy which is enacted in one country with rule of law and democratic processes we get requests from other governments where it would be more challenging to emulate that or adapt it for very different purposes.

DG: Is using trusted flaggers outsourcing your responsibility? We’ve heard this view voiced to us that by not investing enough in these processes, of takedown, or the teams that expedite takedown, that Google and other companies are facilitating intimidation. Is that something you recognise? Is that something you need to increase responsibility for?

KOD: I think this is probably two separate questions and we would probably look at that in two separate ways. The first is – do we have the right resources in place to manage the situation. The second is, what’s the value in a non-monetary sense of the trusted flagger programme. In the first question, I think part of that is wrapped up in the scale of what happens online and the content and the way that Youtube operates. YouTube is very much a platform to give everyone access to share video content. We get about 400 hours worth of content uploaded every minute. It’s an incredibly popular and important platform and adds a great deal of value - whether that’s economic, which it does, but also societal value whether that’s education whether that’s culture whether that’s political debate or news. The way it functions is the ability for every individual to be a content publisher. We then absolutely have to have the right resources and we have the responsibility to make sure that content which is uploaded from our platforms which breaks our guidelines doesn’t maintain on our platforms. So we have invested and we do have teams around the world that are very strong both in engineering capacity but also in the review resources to ensure that when content is flagged it is reviewed and if it should come down it does come down. I think the trusted flagger programme is not an addition in terms of capacity, it’s not kind of ‘x number more’ headcount, it’s an addition in terms of understanding and nuance. So for example we have an issue where, sometimes on Youtube, we have seen videos uploaded from local rappers in London neighbourhoods. There is a real challenge understanding where a local rapper in a London neighbourhood becomes somebody who is actually using local vernacular to make accusations or allegations against a foe. One of the things that is really important to us in there is local expertise. We have people who speak a wide range of languages and we ensure we have people who are embedded in communities which are most relevant to Youtube. I think Yasmin’s point about understanding all the different types of communities is certainly true. We also aren’t experts in everything – we certainly aren’t experts in the vernacular of south London slang! That’s where having expertise whether it’s from law enforcement or a trusted flagger is really really helpful. If you look at the issue of suicide – the Samaritans, for example, have really deep, multi-decades worth of experience of where the right balance is between information which is honest and supportive versus encouraging and damaging. We really do value that expertise and that is where the value of the trusted flagger programme comes in. But it’s certainly not a headcount issue.

DG: Thank you. You mentioned earlier that you don’t regularly publish performance data in terms of how many things are flagged to you, how often they are taken down, what percentage of things are taken down, and how quickly they are taken down. Do you keep that data? Do you keep it at a national level? Not that you have to publish it, but does someone in Google know how much hate speech was reported to the Youtube platform during the six weeks of election campaign, for example? Or what percentage of that was taken down?

KOD:. We’re often asked this request. It is something that it is useful for us to consider more. If you look at something like hate speech – if you look at the flagging process - the flagging process is quite light touch because that enables people to quickly flag. So for example we’ll often get asked the question do we know how much hate speech is flagged by people who are flagging on grounds of sexuality or race, we can look at the content and we can understand what that content is. But if I flag content on there, we won’t know that I am a woman, we won’t know my sexuality or my age. With hate speech in particular, we will have a lot of false positive flags, so people flag Justin Bieber for hate speech. So you could see if you looked at the data, different spikes which are not necessarily associated with genuine hate speech.

DG: But for the things that you do take down, you could have potentially have UK specific data within a short timeframe about how quickly it went from reporting to the time it went down?

KOD: I would have to double check if we keep that on a retrospective.

DG: Because the Committee is interested in making sure there’s progress in this area, and the only way to check if we are making progress is to keep data.

LB: I wonder if you could tell me about any working relationship or partnership you have with the police authorities, in terms of from the start when people might report to you or flag that something that might not be illegal, or not quite illegal but nonetheless inappropriate, what guidance they have have and what guidance you provide to users as well?

KOD: We work with the police in different ways. We have a process where if anyone in the UK police force wants to apply for information or data relating to something that they might have seen on Youtube we have a process that they can do that. And that will go through to our relevant colleagues. Obviously almost everything on Youtube is public – so usually there is evidence and content which is available both for victims and the police to be able to see and to share. We also have dialogues and ways of working with the police on particular issues - for example, on gang crime, as I described in the earlier answer is a good way to see where Met Police and we have had a conversation to understand exactly what the issue is, to understand where creative content meets something that might actually be troubling, or troubling but not law-breaking, or when it is law-breaking. That’s been a relationship and a dialogue which has progressed over the last year or so. In terms of when people find content in our platforms that break the law they are able to flag that to us in the usual process but also report it to the police locally. In exceptions of where we have people livestreaming videos, we have a different level of flagging for that if you believe there is imminent harm, then we have a different process - it’s the same process but we have a different stream - so that people can flag a video if they believe there is imminent harm or they feel there is a risk of imminent harm.

LB: Presumably that’s expedited?

KOD: Yes, absolutely. That then goes into a different response stream.

LB: Do you offer any guidance to people about what they should do? Or is it just a question of flagging to Google if they see something?

KOD: No, I don’t think we do offer specific guidance if people think there is a legal issue there. We have our own community guidelines. And then people are there to understand and interpret the law. That’s something that’s really interesting where society’s, the general population’s understanding is and the police’s understanding is. The MOPAC process is very interesting and very rewarding as actually you have a police force who are thinking very deeply about how it can use existing criminal powers to prosecute on hate crime grounds. And we looked, out of interest, I think in the UK there are about thirteen different laws that can be used to prosecute online hate crime - though that’s a very imprecise number.

And I haven’t had these discussions necessarily with politicians about hate crime but for example with victims of domestic violence or online abuse, often there is a great deal of difference in the interpretation of what happens online. And actually there is very strong due legal process that can be followed by the police and supported by CPS and the law is already there. But actually the cultural understanding within individuals and within the police force and other parts of the judiciary can be improved, I think. And it is very heartening to see the CPS be very clear that hate crime online is illegal and to see that reinforced by the Home Office and I think the Home Secretary herself said something recently which is something we strongly support.

LB: So moving on to legislation, Germany has introduced legislation which requires clearly illegal material to be taken down within 24 hours of it being reported. What would Google’s response be if the government were thinking about introducing regulatory legislation in this country?

KOD: I don’t know the German law closely at all beyond the headlines so I won’t comment on that, unless Dave wants to add anything. We very clearly don’t want hate speech content on our platforms, it’s not what we provide Youtube for and we want to be very effective at keeping that off. I think I would want to understand very specifically how the legislation would work, and how it would get the right balance between hate speech and freedom of speech. Some of that comes back to the discussion we had earlier about the broader international context of how often the UK is looked to as a legislative template for the rest of the world, rightly or wrongly, it just happens to be. When we have seen legislation in relatively related areas, not exactly, we also see interpretations of those used in more stringent ways. So I think I would want to know that we were failing to do this properly through our own terms and conditions and guidelines, which I know that we have and know that we apply. And I would want to see a real body of evidence that showed the scale of the harm and the need for legislation beyond this. But I want us - and Google and Youtube want us - to have a platform which is free from hate speech and free from illegal material.

DS: I would absolutely echo that from Katie. I would also say in terms of the German legislation there is also a concern about the impact on a start-up economy and creating barriers to entry or insurmountable obstacles for small, start-up players, if we were thinking about introducing legislation. I think that is something that we do need to think about when we are talking about the health of the UK tech economy as well. And the point that Katie made very well is that we need to be quite clear that we are fulfilling our terms of service and we are meeting these concerns, before any regulation is considered. That is something that needs to be considered.

SR: We have been looking at the details of different models that have been used and that people talk about. The German model sits under the existing European regulatory framework, so it doesn’t question any of the platform/publisher issue, but it did set a period in which material would have to be taken down if it is clearly illegal, where it’s illegal and that’s obvious even to a non-lawyer. If the UK government are talking about introducing something with a clear expeditious takedown, so you had to meet a certain timeframe, you said if we were going to talk about that you would need to see a body of evidence. But you’ve said yourself, you’re not sure if you collect the data and you certainly don’t publish the data on takedown periods. We’re in a bit of a circle, that it’s quite difficult to collect a body of evidence if there’s no evidence out there about how, for example, Youtube is taking things down within a timeframe.

KOD: There would be public examples that were available. And I think your point about something that is clearly illegal to a non-lawyer - life is not as straightforward as that and some of the context and comments are not as straightforward as that. I think that’s another thing that we are very conscious of. We have our own community guidelines that we adhere to, we’ve written those and it’s absolutely our responsibility to uphold those. We also have the body of UK law which we have a good understanding of and we do have lawyers. I think making case-by-case snapshot decisions on a very short timescale in a highly charged election period is quite a challenge for a private company to do and to ensure there is an ability to have freedom of speech as well. So I think all of those things have to really be considered carefully. The level of harm that we are looking at needs to be carefully weighed against the ability have a diverse conversation which enables freedom of speech. Absolutely there will have been comments on Youtube that break our guidelines that should be removed and have been removed, and very challenging cases that happened in the middle of an election campaign which is very serious and we will work very, very hard to do everything that we commit to. But there’s also this amazing ability to have free dialogue, free expression, free conversation, citizen journalism, angles of election campaigns that weren’t covered in the past, that just weren’t exposed in the past. Youtube is part of that ability to encourage democracy and encourage the debate, and we need to balance that very carefully against an onerous system that makes private companies err on the side of caution and stifles the debate and the democracy that we all want to encourage. We want to enable the best Parliamentary process that we can have. I don’t in any way diminish the difficulty of this issue, and the seriousness of this issue, and the relevance of the issue, but we also have to set it against what we are also trying to maintain.

LB: I take your point. There is this huge tension between absolute free speech and the responsibilities that come with that. But it does seem to me that if you are not publishing the data, if you are not sharing the data, then it’s all on Google’s terms. That Google is deciding whether and what the seriousness of the crimes are, whether they are meeting your guidelines, the guidelines that you set. How is it possible to decide whether legislation is needed if the rest of the world can’t see the evidence that you have?

KOD: I wonder if, if there were cases that were relevant they would help inform that discussion. Your example earlier about a video that was online, that was taken down but taken down after the election is a really interesting case of that. Those conversations and those pieces of evidence can help inform that conversation too.

LB: That leads on to my final question: I think social media companies do accept that things have moved on, and that they need to take more responsibility of what’s going out on their platforms. Would Google be responsive to coming together with other government, with politicians, with other social media companies to talk this through more and see what the way forward might be?

KOD: Absolutely. We would love to do that cognisant of all the other conversations that are already happening and that have been happening. So for example, we’ve been board members of UKCCIS – the government’s online safety council - which started on child online safety but these issues are on a continuum. That’s being reformatted and moving away from child internet safety to everyone's internet safety, and a big part of that is about online abuse, especially of vulnerable groups, women, ethnic minorities, people of certain sexualities. The government itself has its own internet safety strategy which was launched about two or three weeks ago. I think we would be very happy to continue those conversations. We have our own terms and conditions; it’s not in our interest to see those broken, we have them because we want them to be adhered to. I think my one note of caution would be that point about remembering the value of freedom of speech and the ability to exchange information and ideas. We know politicians have a tough time online, and I’m very sympathetic to that, and I know that they can have a tough line and then that crosses over into another line which is completely unacceptable in some cases. But we also know that politics, for time immemorial, has been a very, very robust place to be – there are examples of in-person challenging behaviour and activity. Some of what we discussing is how we, as a greater society, maintain civility whilst at the same time having very robust and important discussions. The value in that point is understanding that we want to preserve the openness of the dialogue, the ability to see how their MPs voted, to see their local MP at a bus stop, to be able to post a video about the rubbish that has been dumped at the end of the road that hasn’t been moved, and to have all the value of democracy in a way that we didn’t have twenty or thirty years ago, but do it in a civil tone. But remember that we are operating within the parameters of real life as well.

DS: There is one thing that I didn’t raise last week that might be worth raising this time. We have launched advanced protection for users at particular risk of hacking et cetera including politicians. I can send you more information and it may not be of direct relevance but it might be worth raising. Basically we give advanced protections to people who are at great risk of hacking, phishing et cetera. There are three aspects to this to further increase the defence on the account. The first is the strongest possible defence against phishing, by requiring security keys to access the account which gives a greater level of protection; we help to protect sensitive data from accidental sharing; and we also help to block fraudulent account access. It’s part of a suite of measures we launched for individuals who might be particularly sensitive to hacking to make sure their accounts remain safe. It’s very recent and I’d be happy to share more information about this with the Committee.

KOD: Can I ask next steps about the report?

LB: We are planning to publish before Christmas, in the form of a Command Paper, it will be presented in Parliament in the Prime Minister’s name.

CSPL Secretariat

November 2017

1. This footnote was subsequently provided by Google: *Jigsaw is an Alphabet company that builds tools to make the world safer. The Jigsaw team tackles a range of global security challenges including defending against digital attacks, mitigating the rise of online hate and harassment, countering online extremism, and fighting censorship. Jigsaw is an Alphabet company, but is independent of Google. The processes described are those of Jigsaw, rather than Google as a whole.* [↑](#footnote-ref-1)
2. In a subsequent email, Google added: *We employ thousands of people at Google and YouTube to review and take action on videos identified by our systems or flagged by our users, work that in turn helps train and develop our machine learning systems to more accurately identify problematic content.*

*These numbers may change over time as we both hire more people and as our machine learning systems get better at identifying content that violates our policies*. *As you can see in our latest YouTube blog on extremism, machine learning is extremely important for flagging violent extremism content for review: “Over 83 percent of the videos we removed for violent extremism in the last month were taken down before receiving a single human flag, up 8 percentage points since August.” -.*

*I hope this information helps clarify our approach to extremism on YouTube. It’s the combination of people and machines that helps us review content at such a large scale when we have over 400 hours of video uploaded every minute and receive over 260,000 user flags a day. We will continue to invest in technology and our teams to improve our review at scale with increased speed and accuracy. As we do, we will continue to publish updates and improvements and discuss them in our conversations with government and civil society.* [↑](#footnote-ref-2)
3. In a subsequent email, Google added: *We remove content when we receive valid legal requests. We report on these requests in our transparency report.* [↑](#footnote-ref-3)