Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act

Ben Wagner
Vienna University of Economics and Business
ben@benwagner.org

Krisztina Rozgonyi
University of Vienna

Marie-Therese Sekwenz
Vienna University of Economics and Business

Jennifer Cobbe
University of Cambridge

Jatinder Singh
University of Cambridge

ABSTRACT

Regulatory regimes designed to ensure transparency often struggle to ensure that transparency is meaningful in practice. This challenge is particularly great when coupled with the widespread usage of dark patterns — design techniques used to manipulate users. The following article analyses the implementation of the transparency provisions of the German Network Enforcement Act (NetzDG) by Facebook and Twitter, as well as the consequences of these implementations for the effective regulation of online platforms. This question of effective regulation is particularly salient, due to an enforcement action in 2019 by Germany's Federal Office of Justice (BfJ) against Facebook for what the BfJ claim were insufficient compliance with transparency requirements, under NetzDG.

This article provides an overview of the transparency requirements of NetzDG and contrasts these with the transparency requirements of other relevant regulations. It will then discuss how transparency concerns not only providing data, but also how the visibility of the data that is made transparent is managed, by deciding how the data is provided and is framed. We will then provide an empirical analysis of the design choices made by Facebook and Twitter, to assess the ways in which their implementations differ. The consequences of these two divergent implementations on interface design and user behaviour are then discussed, through a comparison of the transparency reports and reporting mechanisms used by Facebook and Twitter. As a next step, we will discuss the BfJ's consideration of the design of Facebook's content reporting mechanisms, and what this reveals about their respective interpretations of NetzDG's scope. Finally, in recognising that this situation is one in which a regulator is considering design as part of their action – we develop a wider argument on the potential for regulatory enforcement around dark patterns, and design practices more generally, for which this case is an early, indicative example.

FAT '20, January 27–30, 2020, Barcelona, Spain*

ACM ISBN 978-1-4503-6936-7/20/02…$15.00
https://doi.org/10.1145/3351095.3372856

1 INTRODUCTION

There are numerous challenges with ensuring effective transparency in a socio-technical context. In practice, transparency undertakings are often lacking, with many actors preferring to create the illusion of transparency rather than actually engaging in transparent practices [20]. Notably, even though legal requirements for transparency are common [17, 54], it is relatively uncommon for regulators to enforce these transparency requirements systematically [9, 43]. This makes the fine issued by the BfJ against Facebook "for violating the provisions of the Network Enforcement Act" (NetzDG) [2] particularly interesting. The core job of regulators is to oversee private sector actors and in some cases like the GDPR to oversee public sector actors as well. The issuance of the fine is an evidence that the BfJ was taking on this oversight role and pushing for greater transparency. At the same time, online platforms have a commercial interest in handling complaints using their own Community Standards mechanisms rather than by the provisions of NetzDG, as handling complaints using their Community Standards is considerably easier and cheaper for them [31, 36, 57]. This leads to our main research question: How do Facebook and Twitter implement the transparency provisions of NetzDG and what are the consequences of their respective implementations for the regulation of online platforms?

Platform Community Standards have been heavily criticised as lacking transparency, accountability, and procedural safeguards for the human beings affected by them [1, 13, 14, 19, 30, 64, 65, 68]. Regulation of platforms through instruments like NetzDG can thus be seen as a response to this criticism. The BfJ's decision also needs to be seen in the context of a wider push for platform regulation in Europe; the UK Government's 'Online Harms White Paper'[21] and the French proposal on making social media platforms more accountable [26] are two further examples. The way in which NetzDG is interpreted provides a key indicator of what the future European platform regulation could look like.

1.1 Definitions, Scope and Case Selection

In order to answer the research question, we will first define a few key terms to ensure their meaning is clear within this article, as well as clarify the scope of our analysis. The term *complaints* will be used extensively, as this is the official legal terminology used by the German NetzDG and the BfJ. When speaking about the technical mechanisms used to receive these complaints, we will use the more common terminology of *reporting mechanism* for users to be able to report problematic content. In the context of this article, the terms

complaint and *report* and *flag* can be understood interchangeably as they all refer to a way in which users can inform platforms about problematic content.

We use *dark patterns* to refer to the"instances where designers use their knowledge of human behavior (e.g., psychology) and the desires of end users to implement deceptive functionality that is not in the user's best interest" [27]. §5 explores the concept in detail.

As mentioned in the introduction, this article discusses the transparency provisions of NetzDG. We do not discuss the numerous criticisms of NetzDG in regards to freedom of expressions, which we note have already been extensively discussed elsewhere [16, 33, 35, 37, 38, 69]. Instead, we will focus on the transparency-related aspects of NetzDG, which have not yet received the same level of scrutiny and analysis, and importantly, forms the basis for an enforcement action regarding the means by which online transparency mechanisms are implemented.

In regards to the concept of transparency, we draw on the work of *Flyverbom* to "conceptualize transparency projects as a form of visibility management with extensive and often paradoxical implications for the organizations and actors involved" [20]. From this perspective, it is important to explore "(a) the technological and mediated foundations of transparency and (b) the dynamics of visibility practices involved in efforts to make people, objects, and processes knowable and governable" [20]. As such, this article considers both the technical foundations, the legal and regulatory context, and the attempts by different private sector actors to create visibility through specific forms of compliance with transparency obligations.

As this article includes an empirical analysis, we believe it is important to justify why specific empirical cases were chosen. We decided to analyse how Facebook and Twitter attempt to comply with NetzDG in greater detail. The choice of Facebook was clear, not only as it was the subject of a BfJ enforcement action, but also because it is one of the only platforms we are aware of that exhibits two separate reporting mechanisms in its implementation of NetzDG and platform Community Standards – one of the main reasons it was targeted by the BfJ. Twitter was selected, as it provides a good example of an implementation of NetzDG that combines the reporting mechanisms for Community Standards and NetzDG into a single reporting mechanism. As such, the choice of Facebook and Twitter as cases represents a 'most-different case' [24] selection, allowing for a systematic comparison of the two approaches to implementing NetzDG.

Finally, we contrast the transparency provisions around the content reporting mechanisms provided under NetzDG, with existing mechanisms for transparency and content reporting already provided by online platforms such as Facebook or Twitter. These mechanisms are referred to here as *Community Standards* and consist of all relevant documents and processes which influence how platforms decide which content to remove. For Facebook, this includes the Facebook Community Standards and Terms of Service,[1] as well as internal manuals on how to implement these rules on an everyday basis [4, 67]. For Twitter, we consider the Twitter Rules and policies, general guidelines and Terms of Service.[2]

2 REGULATING TRANSPARENCY

2.1 Transparency requirements in NetzDG

NetzDG came into force in Germany on 1 January 2018 [31]. Its main purpose is to reduce illegal content online through ensuring that platforms block and delete illegal content, while increasing transparency and accountability of platform content removals. According to Medienanstalt Hamburg, NetzDG has led to effective deletion procedures for reported content [29]. NetzDG applies to for-profit media service providers with at least two million registered users that operate online platforms with user generated content [16]. Platforms that provide journalisitic or editorial content do not fall under the scope of NetzDG, nor do instant messaging services like WhatsApp, Wire or Telegram [16].

Since NetzDG was passed, debates about the law have shifted considerably from the context in which it was originally discussed. What was originally debated as a means to increase the compliance with the law by online platforms in Germany increasingly became focused on platform Community Standards, transparency requirements and procedural safeguards for users [57, 61, 69]. For those platforms which fall within the scope of NetzDG, §2 NetzDG specifies that platforms that receive more than 100 notifications about unlawful content per year must publish a public transparency report in German every 6 months. These reports provide an insight into statistics and information necessary for an assessment of NetzDG implementation [31]. Furthermore, they have to provide information about how the platform deals with the reported content it receives.

The minimum reporting standards for these transparency reports are defined in §2 (2) NetzDG. These requirements include a general outline of how criminal activity on the platform is dealt with (Nr.1), a description of the mechanisms in place (Nr. 2), the number of the complaints (Nr. 3), organisational and human resources (Nr. 4), membership of industry bodies (Nr. 5), number of complaints for which an external body was consulted (Nr. 6), number of complaints that were deleted (Nr. 7), time span of deletion or blocking procedure in place (Nr. 8) and measures to inform the user who submitted the complaint, as well as the users whose content is under investigation (Nr. 9). §3 NetzDG also clearly defines how platforms are expected to deal with the complaints of users. §3 (1) NetzDG specified that "the provider of a social network shall maintain an effective and transparent procedure for handling complaints about unlawful content" and that "[t]he provider shall supply users with an easily recognisable, directly accessible and permanently available procedure for submitting complaints about unlawful content"

Finally, §4 NetzDG defines the provisions and regulatory fines that may be imposed. Both intentional or negligent offences fall within the scope of §4 NetzDG. If the provisions of NetzDG are not followed, the fines can reach up to 50 million Euros [31]. Fines can only be imposed for systematic failures of the platforms because of

[1] See: https://www.facebook.com/communitystandards/ and https://www.facebook.com/terms.php

[2] See https://help.twitter.com/en/rules-and-policies/twitter-rules and https://twitter.com/en/tos

mismanagement of their complaint reporting practices [57]. Platforms thus do not violate the rules if they make an 'honest mistake', in the reporting procedures or if content has been overlooked due to an error [16].

2.2 Comparing NetzDG and other transparency regimes

Providing transparency is a crucial element of good regulation. The Organisation for Economic Cooperation and Development (OECD) defined transparency "as one of the central pillars of effective regulation, supporting accountability, sustaining confidence in the legal environment, making regulations more secure and accessible, less influenced by special interests, and therefore more open to competition, trade and investment" [50]. Moreover, transparency contributes to and enables accountability [7, 45] enhancing the overall legitimacy of regulatory decisions [15]. Transparency requirements are often mainstreamed as regulatory tools, typically requiring private actors such as corporations to provide the public with factual and comparable information about their products and practices for various public policy purposes [23]. Organisations like Ranking Digital Rights or the Global Network Initiative also encourage transparency by private companies as a means to ensure greater accountability [6, 42].

Importantly, even if private companies are involved in creating transparency, it is still the responsibility of states to ensure legal certainty and predictability in how transparency is provided [39]. This can be achieved by setting minimum standards for the provision of transparency, or through the regulator providing additional guidance on correct the implementation of transparency provisions. At the same time, corporations often offer the illusion of transparency, providing company information in a way that follow organizational goals at least as much as legal principles [20].

The reporting obligations of §2 NetzDG are most similar to the transparent reporting requirements of telecommunication providers. In the telecommunication sector, transparency regulations are common and used to foster consumer protection policies, focusing on pricing, billing and the quality of service [32]. Furthermore, electronic communication ex ante competition regulation[3] foresees the use of transparent accounting separation and cost accounting systems by the service providers in order to ensure comparable and controllable pricing to be overseen by national regulators. In some cases, telecommunications transparency reports function as corporate responsibility action in response to government surveillance activities [53]. In a similar vein, the "Transparency Reporting Index" by AccessNow[4] compiles information on transparency policies activated as safeguards against government abuses with regards to online surveillance, network disruptions or content removal.

The mere provision of transparency is not a panacea to regulatory concerns but should involve complex mediation processes and the interpretation of information. Therefore, the implementation of transparency requirements should make published data understandable to consumers and to the general public. This was the case, when the German Federal Network Agency (BNetzA) issued in 2016 the Transparency Ordinance for Telecommunications (TKTransparenzV)[5] to improve the information rights of end-users vis-á-vis their service provider with "the provision of transparent, comparable and up-to-date information in a clear, comprehensible and easily accessible form".[6] TKTransparenzV §1 stipulates in detail the information that the product sheets must contain — such as data transmission rates and fundamental contractual terms — and BNetzA further published sample sheets to ensure consistency in the form of information provision. The sample sheets[7] were designed in a simple and comprehensible manner to assist service providers in following transparency rules, while users benefited from information served in an accessible form. However, regarding NetzDG, the BfJ have not provided additional guidance to platform operators on how to design their reporting practices.

2.3 Transparency as visibility management

Companies also use transparency mechanisms as a form of visibility management. Companies like Facebook and Twitter have a strong interest in promoting their perspective on the status and health of their social networks, for which transparency mechanisms are important tools. As noted by *Flyverbom*, organisational positioning is a key way in which transparency is used by firms to "foreclose and downplay other types of intervention that could be made to address the problem and create awareness of how the Internet is governed by both state and non-state actors" [12, 20]. The following section does not claim any specific knowledge of the actual motivations of Facebook or Twitter. Rather, it interprets the different strategies and narratives used by each organisation to justify their decisions in how to comply with NetzDG.

In doing so, Twitter and Facebook engage in very different narratives. Through its combined reporting procedures which integrate NetzDG and Community Standards, Twitter demonstrates that it complies with the requirements of NetzDG to the fullest extent. In order to do this, Twitter lists a large number of cases in its NetzDG transparency reports to show that it is complying with the law wherever possible.

By contrast, the approach taken by Facebook is very different to Twitter. By assessing the vast majority of cases under its own Community Standards, it effectively shows that NetzDG is unnecessary, as its own Community Standards are already highly effective. Following this narrative, it is important to emphasize the low levels of content which is not yet covered by Facebook Community Standards and which would require additional reporting under NetzDG. This approach by Facebook is also consistent with the design of the user interface, which is particularly hard to find and will be discussed in detail in §3.

In this sense, Facebook's NetzDG transparency reporting mechanisms can be seen as a way to contest the legitimacy and necessity of the law. Twitter's interface choices in its reporting mechanisms—by contrast—are consistent with Twitter demonstrating willingness to comply with local legal requirements. Further, there are also

[3] See Commission Recommendation of 19 September 2005 on accounting separation and cost accounting systems under the regulatory framework for electronic communications. OJ L 266, 11.10.2005, p. 64-69.
[4] See https://www.accessnow.org/transparency-reporting-index/.
[5] Verordnung zur Förderung der Transparenz auf dem Telekommunikationsmarkt (TK-Transparenzverordnung - TKTransparenzV) Vom 19. Dezember 2016 (Bundesgesetzblatt Jahrgang 2016 Teil I Nr. 62, ausgegeben zu Bonn am 22. Dezember 2016).
[6] §45n Telekommunikationsgesetz 2003 - TKG 2003.
[7] Product information sheet pursuant to section 1 TKTransparenzV.

clear economic reasons for processing complaints under platform Community Standards rather than under NetzDG [31, 57].

Determining whether a complaint falls under NetzDG following the minimum standards for complaint management provided by NetzDG is more complex and time-consuming than analysing the same piece of content under the platforms' own Community Standards [31, 57]. This is because platform Community Standards are optimised [52] to allow a swift low-cost analysis with no minimum standards [66], while NetzDG aims for legal compliance and includes minimum standards which are more expensive to implement. A similar claim is made by *Keller*, who argues that "maintaining a single set of standards—and perhaps expanding them to accommodate national legal pressure as needed—is much easier" [36].

Transparency requirements are not just being portrayed from a specific perspective by Twitter and Facebook. The German regulator BfJ also has an interest in portraying its implementation of NetzDG in a certain light. The NetzDG has been heavily criticised for creating legal mechanisms that are likely to harm freedom of expression by both German and international legal scholars [57, 61] as well as the UN Special Rapporteur on Freedom of Expression [35] and been the subject of parliamentary hearings to discuss potential revisions [38]. In this political context, showing the effectiveness of the law and its regulatory impact is particularly important.

This political dimension of NetzDG is particularly problematic because BfJ is not an independent regulator, but "has a crucial role in the enforcement of [NetzDG] and directly reports to the Minister of Justice, making it by no means politically independent" [57].

This is one important design flaw in NetzDG, as enforcement was not given to an independent regulator. As such, BfJ may be more inclined to take an adversarial approach to online platforms, in order to demonstrate its ability to regulate key platforms like Facebook. It is thus important to consider the political dimensions of BfJ decisions, where considerations of political visibility are greater than they would be for an independent regulator.

Finally, in line with *Flyverbom* we argue that "transparency efforts, like other visibility practices, always involve selectivity, directionality, and interpretation" [20]. All forms of transparency reporting involve a degree of interpretation. Both companies necessarily need to select which types of content to show and how to interpret the results in the same way that the BfJ needs to do so in determining whether to fine them or not.

As strange as this may seem, the transparency requirements of NetzDG are not just designed to promote transparency. The overall goal pursued by NetzDG is to increase the enforcement of German law against global platforms, and its transparency reporting requirements constitute a mechanism to push for practices which promote the primacy of German law over the Community Standards of platforms. This claim of the primacy of German law over Community Standards is implicit in the BfJ's interpretation of NetzDG in the fine issued to Facebook. As such, how Facebook and Twitter provide transparency has both an organisational and performative dimension that extends far beyond the mere provision of data [3, 20]. The extent to which platforms are willing to adapt their existing organisational practices to acknowledge the primacy of German law over their own Community Standards in their reporting mechanisms is a similarly important factor to the quality of data they provide.

3 IMPLEMENTATIONS OF NETZDG BY FACEBOOK AND TWITTER

3.1 Methodology

In preparing this article we analysed in detail how Facebook and Twitter attempt to comply with NetzDG. It is notable that although a variety of different implementations exist, most providers that we looked at chose to integrate the requirements of the NetzDG directly into their existing user reporting interface. Facebook's approach to NetzDG is different from that of most other platforms, in that they developed a completely separate reporting procedure for complaints under the NetzDG.

Our analysis is based on attempting to flag a fictitious piece of content on 6 August 2019. This piece of fictitious content would, if it were real, constitute incitement to violence and is commonly considered illegal in most jurisdictions as well as in international legal standards. It is also a relatively 'clear case' in regard to the complexity of reporting mechanisms and other types of content such as 'hate speech' would have required additional input.

Before going into detail regarding the individual NetzDG implementations for Facebook and Twitter, we have first provided an overview of the content reporting interfaces used by both platforms for user complaints. As these complaints can be made both based on Community Standards and NetzDG, we have displayed both processes separately for each of the two platforms. The result is an overview of the relevant steps for four different interfaces in total. As Twitter has integrated its NetzDG process into its existing interface for user complaints, the difference in the process involved between lodging a complaint based on NetzDG and Twitter Community Standards is very small. By contrast, there is a significant difference in Facebook's processes for flagging content via the NetzDG and Facebook's interface for content removal via their own Community Standards. Fig.1 provides a visualisation of the steps involved in all four options. We also listed the number of options a user of each interface can choose from, providing an overview of the number of the choices the user is being asked to make.

In order to develop an indicative estimate of how long the different reporting mechanisms would take to complete, we developed a short online survey with identical wording to that used by Facebook and Twitter's reporting mechanisms. This survey was filled out by 190 users at a German-language university in Europe to provide an estimate of how long (in seconds) it would take an average user to read an fill out a form of this kind. The average amount of time it took participants to answer each question is listed in Fig.1.

SMOG grading was used to provide a general overview of readability of the text used by Facebook and Twitter in their reporting mechanisms. SMOG grading has been in use for many decades [44] and remains one of the "fastest and simplest ways to arrive at an objective difficulty rating for reading material" [41], which is also used to analyse the readability of terms and conditions [46] and Internet content more broadly [25]. Fig.1 provides an individual SMOG grade for each step of the process, as well as a grade for all of the text provided by each reporting mechanism. This is to assist in the assessment of the readability of the information being provided to the user. Existing literature on survey design suggest that increasing the length, number of questions and complexity in an online survey is likely to reduce rate at which individuals respond

[18, 28, 46]. Thus it is a plausible hypothesis that longer reporting mechanisms with more options and less readable text are less likely to be completed by the users of online platforms. Additional usage of dark patterns, making the reporting mechanism difficult to find, warning about the negative consequences of using it or redirecting attention away from it, may also result in lower completion rates [27, 43].

3.2 Results

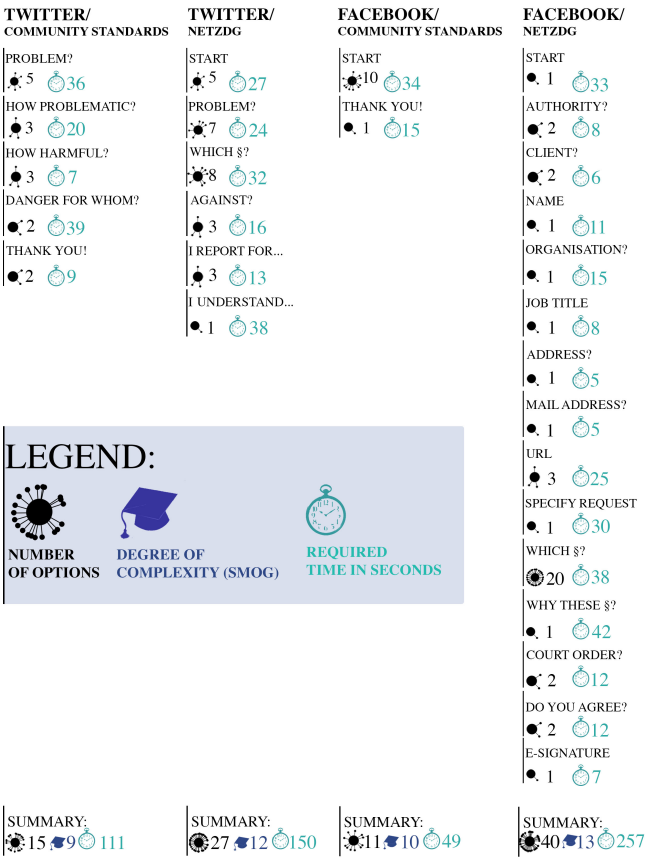

Figure 1: Overview of Facebook and Twitter reporting mechanisms for NetzDG and Community Standards according to number of options, time to answer and complexity of text.

The results of our analysis suggest significant differences for users who want to report content between how Facebook and Twitter implement NetzDG. While there is an additional regulatory 'burden' involved in collecting additional information under NetzDG, Twitter implements this in a manner that does not significantly extend the existing reporting process. The number of questions asked by Twitter expands only slightly from five to six, however the number of answer options increases more significantly from 15 to 27. By contrast, Facebook's Community Standards reporting interface was even shorter, including only two steps; its NetzDG reporting mechanism significantly increases from 2 to 15 steps in total. The number of answer options also increases from 11 to 40. Importantly, the Facebook reporting procedure requires a higher degree of technical expertise from users who want to submit a complaint. Users also need to understand the concept of a URL and submit this URL to Facebook within the form. While Facebook does attempt to explain how to do this to users of the form, the explanation itself may dissuade users from filling in information.

Another important distinction between the implementation of NetzDG by Facebook and Twitter, is that Facebook uses a flat information structure of one long form, while Twitter uses a hierarchical information structure. This means that Facebook ends up showing users lots more information (and thus burdens users with more complexity) than is actually necessary, while Twitter first asks about the concern, then guides the user through the process by narrowing the selection of potentially relevant legal foundations in line with the German law.

3.3 Facebook content reporting interfaces

There is a considerable difference between the processes for reporting complaints under its Community Standards and under the NetzDG on Facebook. To file a complaint under the NetzDG on Facebook, the reporting procedure does not start next to the content the user wants to flag like it does for Community Standards reports, but rather in the Facebook "Help Center". Thus in order to discover this process, the user needs to be aware that:

(1) A German act called NetzDG exists through which they can flag content on Facebook.
(2) That flagging content under NetzDG requires a distinct reporting process on Facebook than the normal Community Standards reporting process.
(3) This process can be found in the Facebook Help Center or in the Impressum of the Facebook website.

In addition, a link to the information page about NetzDG is also provided as part of the legally required basic Imprint/Impressum information about each webpage provided at the bottom of the page [31]. The Impressum used by Facebook is 'hidden' in a similar manner to Terms of Service, Privacy Policy and other documents that users would often never bother to properly read [60]. This approach to compliance by online platforms is not unusual. Many online platforms push information about legal compliance on data protection and their own terms of service as far away from the user as possible [55]. While this behaviour may be common, it can nevertheless be considered a dark pattern, enhancing information asymmetries [43], thereby limiting their ability to access legal rights. Facebook also repeatedly warns users of the dangers of providing information as it "may be required to provide the contents of your report, including your personal information, to parties that have obtained an appropriate court order or as part of other legal process"[8] and attempts to redirect users "[t]o report content as a Community Standards violation instead of under NetzDG".[9] We have provided the exact text used by Facebook in Figure 2.

This even goes to the extent of providing misleading legal advice to users. As noted by *Heldt*: "Not only is this additional reporting

[8] See https://www.facebook.com/help/contact/1909333712721103
[9] See https://www.facebook.com/help/contact/1909333712721103

Figure 2: Facebook NetzDG reporting mechanism attempting to redirect users towards using Community Standards.

procedure well hidden, but once a user is presented with the NetzDG complaint form (on Facebook), he or she will be warned that any false statement could be punishable by law (even if this rule does not apply to statements made to private parties)" [31].

These forms of redirection, obstruction and visual interference can also be considered dark patterns [43], as Facebook makes the process of submitting a NetzDG complaint unnecessarily cumbersome while simultaneously trying to redirect user attention away from the NetzDG reporting process. This shows that dark patterns are not just designed to get users to pay for a product or provide access to their data as is commonly the case, they serve to significantly limit users access to their legal rights under NetzDG. It has been noted that Facebook's NetzDG reporting mechanisms can "prevent users from reporting potentially unlawful content, which is cause for concern as it may result in chilling effects" [31].

However it should also be noted that such effects are not limited to Facebook's implementation of NetzDG, but also extend to other reporting mechanisms. There is extensive literature on the chilling effects associated with DMCA counter-notices, which are rarely used due to the threat of liability [8, 58, 62, 63]. Even though users in theory have access to the right to counter-notice, the way DMCA legislation is designed provides considerable disincentives to access this right in practice. Facebook's design choices also lead to chilling effects for users, however these chilling effects are based not on legal provisions but instead on Facebook's design choices.

All reported content is first reviewed under Facebook Community Standards. If these are violated, then the piece of content is removed globally. If the content in question only violates German law but not Facebook's Community Standards, then the content is blocked only in Germany. This two-step approach has allowed Facebook to avoid having its broader content review process subject to NetzDG transparency reports. In some respects, Facebook's community guidelines are stricter than those under NetzDG. For example, Facebook bans types of content such as nudity, while German laws do not have a general prohibition on nudity on social network sites [16]. Facebook's interpretation of the scope of NetzDG is crucial in this context, as they seem to interpret the German NetzDG as a 'legal add-on' to their existing global Community Standards and design their internal processes accordingly. As a result, analysing the content to their own Community Standards takes precedent over ensuring that it complies with German law.

At the same time, this interpretation of NetzDG is convenient for Facebook, as it limits the transparency requirements the company falls under. In this reading of NetzDG, it does not have to list complaints made under the NetzDG if they were already covered by Facebook Community Standards. As a result, the transparency reports listed below include a very low number of NetzDG complaints. As comparable statistics for Facebook Community Standards violations in Germany are not available, it is impossible from the outside to say with complete certainty how many complaints were made both in the Facebook NetzDG interface and in the Facebook Community Standards interface. As noted above, this very limited interpretation of transparency requirements is not unusual, however the limited scope of interpretation of NetzDG conflicts with both the spirit and wording of the regulation [31].

3.4 Twitter content reporting interfaces

By contrast, Twitter has implemented a far broader understanding of NetzDG in its reporting mechanism. By integrating a NetzDG complaints mechanism as part of its existing reporting mechanism, there are few differences between Twitter's existing Community Standards complaints form and the NetzDG reporting mechanism. To report content on Twitter, the user has to click on the three dots next to the post to open a pop-up window where one can choose between the following options: "I'm not interested in this Tweet", "It's suspicious or spam", "It displays a sensitive image", "It's harmful" or "Covered by Netzwerkdurchsetzungsgesetz". If the user clicks on the last option, they will have to give further information about the problematic substance of the Tweet. Moreover by organising information about the complaint in a hierarchical manner, Twitter significantly reduces the level of complexity a user faces.

Importantly, Twitter also has an organisational process for dealing with NetzDG complaints that is distinct from Facebook: "In order to comply with NetzDG, Twitter depends on a staff of 50 and a separate reporting flow to analyse German complaints. Elsewhere, content is first reviewed against Twitter's terms and conditions. In Germany, it is analysed against the NetzDG's narrower definition of illegal content" [16]. Finally, it is important to mention that Twitter's NetzDG reporting interface is only provided in German, which may be a challenge for Twitter users in Germany that do not speak German.

3.5 How many NetzDG complaints are actually made?

Taking a holistic approach to understanding transparency also requires looking at "interpretation, and negotiation processes; and consequences" [3] that result from attempts at creating transparency. In this section we will look in greater detail at the consequences of the design choices taken by Facebook and Twitter in order to comply with NetzDG.

Because public transparency reports are required by §2 NetzDG, it is also possible to see how these design choices influenced users decisions to submit complaints via the respective platforms under NetzDG. Based on the NetzDG transparency reports provided by Facebook and Twitter, we compiled a summary (Fig.3) of the number of reports made, in which quarter, by which entities (complaints bodies or individual users) and how the platforms responded.

As Facebook has a far larger number of active users in Germany than Twitter, we also provided an additional metric of 'Reports per million users.' Based on aggregating data from several different sources [5, 47, 56] (Facebook: avg. 31,5 Million 2018, 32 Million 2019; Twitter: avg. 3,8 Million 2018 and 2019) we believe it is possible to provide a more accurate estimate of how many reports are provided by individual users.

	Facebook											
	Q1-Q2 2018				Q3-Q4 2018				Q1-Q2 2019			
	By compl aints bodies	Action taken	By users	Action taken	By complai nts bodies	Action taken	By users	Action taken	By complai nts bodies	Action taken	By users	Action taken
Complaints & action	**113**	**45**	**773**	**492**	**92**	**57**	**408**	**320**	**123**	**103**	**551**	**338**
Reports per Million Users*	**0.35**	**0.14**	**2.45**	**1.56**	**0.29**	**0.18**	**1,29**	**1,01**	**3.84**	**3.21**	**17.2**	**10.5**

	Twitter											
	Q1-Q2 2018				Q3-Q4 2018				Q1-Q2 2019			
	By compl aints bodies	Action taken	By users	Action taken	By complai nts bodies	Action taken	By users	Action taken	By complai nts bodies	Action taken	By users	Action taken
Complaints & action	**20754**	**1533**	**244064**	**27112**	**20140**	**1161**	**236322**	**22004**	**26376**	**1950**	**472970**	**44752**
Reports per Million Users*	**5461**	**403**	**64227**	**7134**	**5300**	**305**	**32190**	**5790**	**6941**	**513**	**124465**	**11776**

Figure 3: Facebook and Twitter NetzDG Transparency Report Original and Weighted by Number of Users.

The results of this analysis are stark: user complaints per million users listed in NetzDG transparency reports are between **7236 and 26215 times higher on Twitter than they are on Facebook**. There are several possible explanations for this massive discrepancy of reporting NetzDG complaints between Facebook and Twitter:

(1) Facebook is first checking NetzDG reports under its Community Standards reporting mechanisms and if these reports fall under their own Community Standards, not reporting them as part of their transparency reports. Twitter is not.
(2) The NetzDG reporting mechanism is much harder for users to find on Facebook than on Twitter. As a result Facebook users are less likely to make reports under NetzDG than Twitter users.
(3) The Facebook NetzDG reporting mechanism is designed in a way that users are unlikely to complete it. This could lead to a considerable number of users not completing reports made under this form. Twitter does not do this.

This leads to far higher numbers of reported complaints for Twitter NetzDG transparency reports than for Facebook NetzDG transparency reports. Without more data on the the actual usage of the Community Standards and NetzDG reporting mechanisms, it is impossible to say which of these three factors has the greatest influence. What seems evident is that the design choices made by both Twitter and Facebook influence to a considerable degree the number of complaints they receive under the NetzDG.

An alternate hypothesis could be that Twitter has much more illegal content on its platform, or that German Twitter users are far more likely to complain about content in general. However it seems unlikely that either of these two hypothesis would result in differences at such a scale.

In this context, what Facebook and Twitter consider a NetzDG complaint is particularly important; a NetzDG complaint needs to be following basic minimum standards and comes with additional rights for individuals making complaints under NetzDG. In contrast, complaints under the platforms' Community Standards do not have these protections. By Facebook not considering user complaints filed by their Community Standards reporting mechanisms to be NetzDG complaints, and thus not including them in their transparency reports, Facebook is both hindering oversight while effectively limiting user rights.

4 FINING FACEBOOK FOR VIOLATING NETZDG TRANSPARENCY REQUIREMENTS

In July 2019, the first fine of 2 Million Euros was issued pursuant to the German NetzDG by Germany's BfJ. According to the BfJ Facebook's NetzDG transparency report "fails to meet a number of statutory information requirements" [2]. One of the central problems the report reveals is the dual-reporting system Facebook has in place: "The Federal Office of Justice understands that a considerable number of reports are made via the widely known standard channels [on the basis of Facebook's Community Standards], and that the account given in the published report is therefore incomplete" [2]. The Federal Office of Justice criticizes the transparency report of Facebook because the indicated statistics and numbers are flawed. On the 19th of July 2019, Facebook filed an objection against the fine and the decision by BfJ [51].

The purpose of §2 NetzDG is to ensure that the BfJ is in a position to effectively oversee the efforts undertaken by the platforms to eliminate users' criminally punishable activity and the processes, mechanisms put in place to that. The NetzDG rules and the reporting obligations reflect the concept of procedural accountability [10], whereby regulators investigate platform operators' governance procedures to ascertain whether they have adopted appropriate processes in making, implementing and enforcing their rules. Following this concept of procedural accountability, unified and comparable data and information reported by the platforms is the fundamental tenet of the regulatory oversight. The problem with Facebook's reporting manner is that they hinder regulatory supervision. Once the data on the 'amount' of hate speech or terrorist content does not reflect accurate volumes of such supposedly illegal speech, none of the actions by the platforms can be evaluated. Comparability across platforms is critical to gauge on the actual implementation of NetzDG. Furthermore, the lack of in-depth information on the trends of users' behaviour, as it relates to distinct illegal speech categories, hinders any policy response.

According to the BfJ, Facebook's separate reporting paths therefore violate §2 (1) first sentence in conjunction with §2 (2) Nr. 7 NetzDG, because the numbers of complaints received about unlawful content do not mirror reality: "The incomplete data on the number of complaints also affects the degree to which meaningful, disaggregated information has been provided on the measures taken in response [...] The incomplete figures on complaints mean that an account of how effective the complaints mechanism is, as foreseen by the legislation, is rendered impossible" [2].

4.1 Scope of user complaints covered by NetzDG

One of the main challenges with NetzDG is correctly assessing the scope of the complaints covered by it. The law itself is relatively vague in this regard and does not necessarily limit the scope of user complaints covered by the law, beyond saying that they should constitute complaints about illegal content.

It does however, in §3 (1) NetzDG, specify that providers of social networks should "supply users with an easily recognisable, directly accessible and permanently available procedure for submitting complaints about unlawful content". However, the German BfJ did not argue to fine Facebook because they did not have a complaints procedure, but rather because they violated transparency requirements by providing incomplete transparency reports. This begs the question—in the context of NetzDG—what would be considered a complete transparency report by BfJ in these circumstances?

There are four possible answers to this question. Transparency reports could be considered necessary for:

(1) All complaints made by users through the official NetzDG reporting portal of the online platform;
(2) All complaints made by users that refer to the NetzDG in some way or form, regardless of the ways in which these complaints are made;
(3) All complaints in which users explicitly refer to the illegality of a specific piece of content;
(4) All complaints made by users which are *de facto* about illegal content, regardless of whether the user references the legality or NetzDG.

Importantly the possible answers to this question become increasingly costly and time consuming for the platform, the further down the list they progress. In its most expansive interpretation of NetzDG (#4 above), platforms would have to essentially check all complaints made under any reporting mechanism on their platform twice, once regarding their own Community Standards and once for NetzDG. As there are legal requirements for the minimum qualifications of staff working on NetzDG in contrast to staff working on Community Standards which have no such requirements (in law), it can be assumed that option 4 would be considerably more costly and burdensome than option 1, in particular when combined with the potential need for checking the same content twice to different standards.

So how could providers ensure that they limit the scope of the implementation of NetzDG suggested in option 4? They can most effectively do so by providing upfront and clear information to users at an early stage when a complaint is made, for example like Twitter's content reporting mechanism. By providing this information as part of every complaints process, they are better positioned to claim that they informed the user of their rights and the user did not wish to consider their complaint as part of the NetzDG.

By contrast, Facebook's choice of interface design puts it at risk of a far more extensive interpretation of NetzDG. Even if it were to provide transparency reporting for all complaints made using its NetzDG form (which the German BfJ argues it does not), it could still be expected to provide transparency reporting for all other claims made as part of its Community Standards reporting process. It could be argued that these could also potentially be covered by NetzDG, as users were not sufficiently informed of their ability to claim these rights under the NetzDG and that the scope of NetzDG could thus extend to such claims.

Put in the context of interface design, this means that a well-designed and easy-to-use user-interface that promotes genuine transparency and gives users easy access to their rights reduces the potential scope of NetzDG. This interpretation of the act still needs to be argued in court, but the BfJs line of argumentation already shows promising signals for its ability to reduce the incentives for dark patterns in interface design.

5 NETZDG: REGULATING DESIGN?

A key aspect of the enforcement action of this case relates to the design decisions made by Facebook. This bears exploration, given the BfJ action is an indicative example of potential and means for regulators to contest the usage of dark patterns. In light of this, we now explore the potential for regulatory enforcement around dark patterns and design practices more generally.

5.1 Dark patterns

As mentioned, dark patterns are essentially design practices that seek to unduly influence and manipulate users. Definitions of the term include: "a user interface that has been carefully crafted to trick users into doing things" [9]; "instances where designers use their knowledge of human behavior (e.g., psychology) and the desires

of end users to implement deceptive functionality that is not in the user's best interest" [27]; and "user interfaces whose designers knowingly confuse users, make it difficult for users to express their actual preferences, or manipulate users into taking certain actions" [40].

The concept of dark patterns comes from research on computer-human interaction, which uses the term *patterns* to describe specific types of interaction between an interface and a human being which solve a specific problem [48]. One common example of such design patterns is a 'wizard', which is frequently used to lead users through a specific set of questions to solve a specific problem. As design patterns, wizards all tend to have a similar look and feel and are designed in a similar manner. While dark patterns also share similar design traits, they have developed in recent years into a conceptual framework to describe manipulative design practices in digital technologies which attempt to manipulate human behaviour.

The concept has gained prominence in recent years, and has been included in public policy documents, most notably in a report of the Norwegian Consumer Council on Privacy and Dark Patterns in 2018 [34]. More recently *Mathur et al.* have also been able to show that dark patterns are a common phenomenon, with 11.1 per cent of the "11K most popular shopping websites" [43] using some form of dark pattern.

5.2 Regulating dark patterns

Much of the literature on dark patterns argues that that they are very hard to regulate. *Brignull* suggests that greater user awareness could be helpful [9], while *Gray et al.* believe that "ethical standards are necessary" [27] to respond to the challenge of dark patterns. However as noted by *Mathur et al.*, some of the commonly considered dark patterns "are unambiguously unlawful in the United States (under Section 5 of the Federal Trade Commission Act and similar state laws), and the European Union (under the Unfair Commercial Practices Directive and similar member state laws)" [43]. *Mathur et al.* also argues other dark patterns are "likely in violation of affirmative disclosure and independent consent requirements in the Consumer Rights Directive as well as the General Data Protection Regulation (GDPR)" [43]. Even if dark patterns are considered unlawful, these laws are rarely enforced [9, 11, 43].

One aspect that makes the BfJ decision particularly interesting is that it indicates the potential and means for regulators to contest the usage of dark patterns. Many design decisions which are seen to be in conflict with the NetzDG by the BfJ, would also be considered typical examples of dark patterns, i.e. those designs aimed to divert, obstruct and redirect [9, 27, 43]. Given the prevalence of "unambiguously unlawful" [43] dark patterns online, the NetzDG and the associated enforcement actions of the BfJ[10] represent an example for exploring, in broader terms, how the regulation of dark patterns could be effective.

5.3 Regulating beyond the user interface

A central aspect of the BfJ's action against Facebook concerned the NetzDG's requirement for platforms to employ an "effective and transparent measure" for *individuals* to report content that they consider problematic.

Dark patterns concern user manipulation. It follows that such design practices may be particularly attractive where (i) regulatory requirements require some interaction with a user, be it the provision of information or direct action (e.g. reporting); and (ii) there are incentives for the organisation to limit that interaction. In this example, an interface design that results in users submitting fewer NetzDG reports can bring benefits in terms of a lower administrative burdens, amongst others.

However, many transparency obligations, rather than concerning an individual's interaction with the organisation (and their systems), will fall on the organisation's processes. For instance, provisions of the NetzDG require platforms to provide the regulator with detail on how certain complaints are handled (see §2.1). As such, the design of business processes, workflows, and the technology are important considerations, as is the degree to which these systems support interrogation and oversight by relevant parties, be they internal or external (regulators, users, etc.). These aspects directly impact accountability; concepts such as 'procedural accountability' are predicated on these foundations [10].

Such considerations have implications from the technical design perspective, beyond that of the user interface. That is, there appears opportunities for technical mechanisms that support transparency (and accountability more generally) to be designed and integrated within technical system architectures [59]. These could include, for instance, means facilitating secure audit; recording and verifying data transfers and information sharing; managing automation; etc.

So just as enforcement might work to dissuade dark patterns and problematic interface design, it will be interesting to see whether and how regulatory actions—particularly with the trend towards increased complexity and automation—also work to influence the design, architecture, and other aspects of technical systems.

6 CONCLUSION

How do Facebook and Twitter implement the transparency provisions of NetzDG and what are the consequences of their respective implementations for the regulation of online platforms? This article has shown that there are considerable differences between the implementations of NetzDG by Facebook and Twitter. Both the usage of dark patterns and the way user complaints are organised and structured within Facebook and Twitter have considerable consequences for wider platform governance. This is particularly the case as Facebook was fined for its NetzDG transparency practices by the German BfJ.

But also beyond this individual enforcement action, the Facebook-NetzDG case is of wide-reaching importance. There are many policy and legislative proposals in the making at the moment in Europe, including the UK Government's 'Online Harms White Paper' [21] and the French proposal on making social media platforms more accountable [26]. These foresee the central involvement of national regulators in the implementation of the proposed new rules on platforms. The decision of the BfJ in this context will likely set a precedent for how future cases are decided.

[10] Of course, the BfJ's remit is limited to a particular scope, and so their ability to regulate does not extend to dark patterns in other areas such as privacy or consumer protection.

Mechanisms for transparency are "effective only when the information they produce becomes embedded in everyday decision-making routines of users and disclosers" [22]. The current case implies that these routines should be carefully designed and realized based on mutual and consensual interpretation of the law and the logic on which the law is based. Regulators should assist such practices with the elaboration of reporting guidelines and further direct the platforms in their operations in a consultative manner. Germany is carefully looked at by many right now: the principles of 'good regulation' [49, 50] are being tested in this case.

What NetzDG also demonstrates is that regulatory requirements for transparency around illegal content regulation are insufficient. Indeed many of the topics that NetzDG initially set out to combat like 'hate speech' or 'misleading information online' are not illegal, and are captured under the Community Standards of each specific platform. As the boundaries between these Community Standards and illegal content are typically blurry, a lack of reporting around the moderation of legal content opens the door to precisely the kind of category shifting between NetzDG and Community Standards that Facebook has engaged in. Thus the use of NetzDG to promote better design practices in online social networks are promising. Particularly in the area of safeguarding transparency for users around reporting mechanisms, NetzDG could work to reduce the prevalence of dark patterns and manipulative user design.

Another related challenge is the lack of standardisation in online content moderation. This is particular important, as the reporting requirements for NetzDG are based on user reported data. It is thus of great importance to ensure the consistency of this user reported data, as well as ensure that it is provided in a standardised and comparable manner. As the metrics used differ from platform to platform, this makes meaningful comparisons extremely difficult. NetzDG and other transparency regimes could go much further here, by more precisely defining the exact scope of existing reporting requirements and ensuring that all published data has to be audited. These challenges also apply to user interfaces, as making the same reporting mechanisms easier to use on one platform than another evidently influences how the number of user reports received can be interpreted.

At the core of the debate however remains the question of primacy of standards within content reporting mechanisms. Facebook believes that its Community Standards are to be considered first and only if the do not apply then 'local' regulatory requirements should be considered. While it is understandable that a global social network would want to standardise its content moderation practices around a single set of standards, it appears to prioritise its own 'global' rules over 'local' laws in the process. By designing and operating on the assumption that its Community Standards supersede local laws, Facebook not only sets a global default for speech online [66], it is neither transparent nor accountable in doing so [1, 19, 30, 68].

NetzDG has been criticised extensively for its potential to limit freedom of expression online [16, 33, 35, 37, 38, 69]; however its transparency provisions have received much less attention. Should the BfJ's enforcement actions be upheld in court, NetzDG will represent a practical example of regulatory enforcement influencing design. This serves to indicate what a regulatory enforcement regime that seeks to increase transparency and discourage poor and misleading design could look like. Going forward, this might encourage regulatory and enforcement efforts that incentivise design practices that improve transparency, and more generally, that support broader accountability regimes and human rights online.

REFERENCES

[1] ACLU, EFF, Irina Raicu, Sarah T. Roberts, CDT, Open Technology Institute, Nicolas Suzor, and Sarah Myers West. 2018. Santa Clara Principles on Transparency and Accountability in Content Moderation. https://santaclaraprinciples.org/images/scp-og.png

[2] BfJ Press Agency. 2019. Federal Office of Justice Issues Fine against Facebook. https://www.bundesjustizamt.de/DE/Presse/Archiv/2019/20190702_EN.html;jsessionid=2991FA58C1ACD667E932F1E75AC007FB.2_cid370?nn=3451904

[3] Oana Brindusa Albu and Mikkel Flyverbom. 2019. Organizational Transparency: Conceptualizations, Conditions, and Consequences. *Business & Society* 58, 2 (2019), 268–297. https://doi.org/10.1177/0007650316659851

[4] Julia Angwin and Hannes Grassegger. 2017. Facebooks Secret Censorship Rules Protect White Men from Hate Speech But Not Black Children. https://www.propublica.org/article/facebook-hate-speech-censorship-internal-documents-algorithms

[5] We are social. 2019. Digital 2019 in Germany. https://wearesocial.com/de/digital-2019-germany

[6] DorothÃĺe Baumann-Pauly, Justine Nolan, Auret van Heerden, and Michael Samway. 2017. Industry-Specific Multi-Stakeholder Initiatives That Govern Corporate Human Rights Standards: Legitimacy assessments of the Fair Labor Association and the Global Network Initiative. *Journal of Business Ethics* 143, 4 (2017), 771–787. https://doi.org/10.1007/s10551-016-3076-z

[7] Mark Bovens, Robert E. Goodin, and Thomas Schillemans. 2014. *The Oxford Handbook of Public Accountability.* Oxford University Press, Oxford, UK. https://doi.org/10.1093/oxfordhb/9780199641253.013.0012

[8] Annemarie Bridy and Daphne Keller. 2016. *U.S. Copyright Office Section 512 Study: Comments in Response to Notice of Inquiry.* SSRN Scholarly Paper ID 2757197. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=2757197

[9] Harry Brignull. 2019. Dark Patterns. https://www.darkpatterns.org/

[10] Mark Bunting. 2018. From editorial obligation to procedural accountability: policy approaches to online content in the era of information intermediaries. *Journal of Cyber Policy* 3, 2 (2018), 165–186.

[11] Michael Chromik, Malin Eiband, Sarah Theres Voelkel, and Daniel Buschek. 2019. Dark Patterns of Explainability, Transparency, and User Control for Intelligent Systems. In *Joint Proceedings of the ACM IUI 2019 Workshops co-located with the 24th ACM Conference on Intelligent User Interfaces (ACM IUI 2019), Los Angeles, USA, March 20, 2019.* ACM IUI 2019 Workshop, Los Angeles, USA, 6. http://ceur-ws.org/Vol-2327/IUI19WS-ExSS2019-7.pdf

[12] DeNardis. 2014. *The global war for internet governance.* Yale University Press, New Haven. http://www.worldcat.org/title/global-war-for-internet-governance/oclc/844731628&referer=brief_results

[13] Nicholas Diakopoulos. 2015. Algorithmic Accountability: Journalistic investigation of computational power structures. *Digital Journalism* 3, 3 (2015), 398–415. https://doi.org/10.1080/21670811.2014.976411

[14] Nicholas Diakopoulos. 2016. Accountability in algorithmic decision making. *Commun. ACM* 59, 2 (2016), 56–62.

[15] Susan E. Dudley and Kai Wegrich. 2016. The role of transparency in regulatory governance: Comparing US and EU regulatory systems. *Journal of Risk Research* 19, 9 (2016), 1141–1157. https://doi.org/10.1080/13669877.2015.1071868

[16] William Echikson and Olivia Knodt. 2018. *Germanys NetzDG: A Key Test for Combatting Online Hate.* SSRN Scholarly Paper ID 3300636. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=3300636

[17] Lilian Edwards and Michael Veale. 2018. Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'? *IEEE Security & Privacy* 16, 3 (2018), 46–54.

[18] Philip James Edwards, Ian Roberts, Mike J. Clarke, Carolyn Diguiseppi, Reinhard Wentz, Irene Kwan, Rachel Cooper, Lambert M. Felix, and Sarah Pratap. 2009. Methods to increase response to postal and electronic questionnaires. *The Cochrane Database of Systematic Reviews* 3, 3 (2009), 527. https://doi.org/10.1002/14651858.MR000008.pub4

[19] Christina Fink. 2019. Dangerous Speech, anti-muslim violence and Facebook in Myanmar. *Journal of International Affairs* 71, 1 (2019), 11.

[20] Mikkel Flyverbom. 2016. Digital Age Transparency: Mediation and the Management of Visibilities. *International Journal of Communication* 10, 0 (2016), 13. https://ijoc.org/index.php/ijoc/article/view/4490

[21] UK Department for Culture Media and Sport. 2019. *Online harms white paper.* Technical Report. UK Department for Culture Media and Sport, London, UK. 97 pages. https://www.gov.uk/government/consultations/online-harms-white-paper OCLC: 1105739026.

[22] Archon Fung, Mary Graham, and David Weil. 2007. *Full Disclosure: The Perils and Promise of Transparency.* Cambridge University Press, Cambridge, UK.

[23] Archon Fung, Mary Graham, David Weil, and Elena Fagotto. 2004. *The Political Economy of Transparency: What Makes Disclosure Policies Effective?* Technical Report. Transparency Policy Project, John F. Kennedy School of Government. 56 pages.

[24] John Gerring. 2007. *Case study research : principles and practices.* Cambridge University Press, New York.

[25] Anindya Ghose, Panagiotis G. Ipeirotis, and Beibei Li. 2012. Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content. *Marketing Science* 31, 3 (2012), 493–520.

[26] French Government. 2019. Creating a French framework to make social media platforms more accountable: Acting in France with a European vision. http://thecre.com/RegSM/wp-content/uploads/2019/05/French-Framework-for-Social-Media-Platforms.pdf

[27] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* ACM, New York, NY, USA, 534:1–534:14. https://doi.org/10.1145/3173574.3174108 event-place: Montreal QC, Canada.

[28] Yimeng Guo, Jacek A. Kopec, Jolanda Cibere, Linda C. Li, and Charles H. Goldsmith. 2016. Population Survey Features and Response Rates: A Randomized Experiment. *American Journal of Public Health* 106, 8 (2016), 1422–1426. https://doi.org/10.2105/AJPH.2016.303198

[29] Medienanstalt Hamburg. 2019. MA HSH-Auswertung der Transparenzberichte nach NetzDG. https://www.ma-hsh.de/infothek/publikationen/ma-hsh-auswertung-der-transparenzberichte-nach-netzdg.html Medienanstalt Hamburg.

[30] Marjorie Heins. 2013. The Brave New World of Social Media Censorship. *Harvard Law Review Forum* 127, 8 (2013), 325–330. https://heinonline.org/HOL/P?h=hein.journals/forharoc127&i=328

[31] Amelie Heldt. 2019. Reading between the lines and the numbers: an analysis of the first NetzDG reports. *Internet Policy Review* 8, 2 (2019), 18. https://doi.org/10.14763/2019.2.1398

[32] ITU. 2013. *Regulation and consumer protection in a converging environment.* Technical Report. ITU, Geneva, Switzerland. 32 pages. https://www.itu.int/ITU-D/finance/Studies/consumer_protection.pdf

[33] Katharina Kaesling. 2019. Privatising Law Enforcement in Social Networks: A Comparative Model Analysis. *Erasmus Law Review* 11 (2019), 151–164. https://doi.org/10.5553/ELR.000115

[34] Oyvind H. Kaldestad. 2018. Deceived by design. https://www.forbrukerradet.no/undersokelse/no-undersokelsekategori/deceived-by-design/

[35] David Kaye. 2018. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression.* Technical Report A/HRC/38/35. United Nations, Geneva, Switzerland. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G18/096/72/PDF/G1809672.pdf?OpenElement

[36] Daphne Keller. 2019. *State and platform hybrid power over online speech.* Technical Report. Hoover Institution Essay. 40 pages. https://assets.documentcloud.org/documents/5699593/Who-Do-You-Sue-State-and-Platform-Hybrid-Power.pdf

[37] Matthias C. Kettemann. 2018. The Future of the NetzDG: Balanced Briefing Materials on the German Network Enforcement Act.

[38] Matthias C. Kettemann. 2019. Stellungnahme als Sachverstaendiger fuer die oeffentliche Anhoerung zum Netzwerkdurchsetzungsgesetz auf Einladung des Ausschusses fuer Recht und Verbraucherschutz des Deutschen Bundestags, 15. Mai 2019. https://www.hans-bredow-institut.de/uploads/media/default/cms/media/up8o1iq_NetzDG-Stellungnahme-Kettemann190515.pdf

[39] Aleksandra Kuczerawy. 2018. *Intermediary Liability and Freedom of Expression in the EU: From Concepts to Safeguards.* Intersentia, Cambridge, United Kingdom.

[40] Jamie Luguri and Lior Strahilevitz. 2019. *Shining a Light on Dark Patterns.* SSRN Scholarly Paper ID 3431205. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=3431205

[41] George H. Maginnis. 1982. Easier, Faster, More Reliable Readability Ratings. *Journal of Reading* 25, 6 (1982), 598–599.

[42] Nathalie MarÃĬchal. 2015. Ranking digital rights: Human rights, the Internet and the fifth estate. *International Journal of Communication* 9, 10 (2015), 3440–3449.

[43] Arunesh Mathur, Gunes Acar, Michael Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *arXiv:1907.07032 [cs]* 1, 1 (2019), 32. http://arxiv.org/abs/1907.07032 arXiv: 1907.07032.

[44] G. Harry McLaughlin. 1969. SMOG Grading- a New Readability Formula. *Journal of Reading* 12, 8 (1969), 639–646.

[45] Albert Meijer. 2014. Transparency. In *The Oxford Handbook of Public Accountability.* Oxford University Press, Oxford, UK, 507–524. https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199641253.001.0001/oxfordhb-9780199641253-e-043

[46] Stuart Moran, Ewa Luger, and Tom Rodden. 2014. Literatin: Beyond Awareness of Readability in Terms and Conditions. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication (UbiComp '14 Adjunct).* ACM, New York, NY, USA, 641–646. https://doi.org/10.1145/2638728.2641684 event-place: Seattle, Washington.

[47] NapoleonCat. 2019. Number of Facebook users in Germany. https://www.statista.com/statistics/1017402/facebook-users-germany/

[48] Erik G. Nilsson. 2009. Design patterns for user interface for mobile applications. *Advances in engineering software* 40, 12 (2009), 1318–1328.

[49] OECD. 2005. *Guiding Principles for Regulatory Quality and Performance.* OECD, Paris, France. https://www.oecd.org/fr/reformereg/34976533.pdf

[50] OECD. 2011. *Regulatory Policy and Governance: Supporting Economic Growth and Serving the Public Interest.* OECD, Paris, France. https://doi.org/10.1787/9789264116573-en

[51] Heise Online. 2019. Facebook wehrt sich gegen NetzDG-Bussgeld. https://www.heise.de/newsticker/meldung/Facebook-wehrt-sich-gegen-NetzDG-Bussgeld-4475699.html dpa.

[52] Rebekah Overdorf, Bogdan Kulynych, Ero Balsa, Carmela Troncoso, and Seda Guerses. 2018. POTs: Protective Optimization Technologies. (2018). http://arxiv.org/abs/1806.02711 arXiv: 1806.02711.

[53] Christopher Parsons. 2019. The (In)effectiveness of Voluntarily Produced Transparency Reports. *Business & Society* 58, 1 (2019), 103–131. https://doi.org/10.1177/0007650317717957

[54] Christopher A. Parsons. 2015. *Do Transparency Reports Matter for Public Policy? Evaluating the Effectiveness of Telecommunications Transparency Reports.* SSRN Scholarly Paper ID 2546032. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=2546032

[55] James Pierce, Sarah Fox, Nick Merrill, Richmond Wong, and Carl DiSalvo. 2018. An Interface Without A User: An Exploratory Design Study of Online Privacy Policies and Digital Legalese. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18).* ACM, New York, NY, USA, 1345–1358. https://doi.org/10.1145/3196709.3196818 event-place: Hong Kong, China.

[56] Philipp Roth. 2019. Offizielle Facebook Nutzerzahlen fuer Deutschland (Stand: Maerz 2019). https://allfacebook.de/zahlen_fakten/offiziell-facebook-nutzerzahlen-deutschland

[57] Wolfgang Schulz. 2018. *Regulating Intermediaries to Protect Privacy Online.* SSRN Scholarly Paper ID 3216572. Social Science Research Network, Rochester, NY. https://papers.ssrn.com/abstract=3216572

[58] Wendy Seltzer. 2010. Free speech unmoored in copyright's safe harbor: Chilling effects of the DMCA on the first amendment. *Harv. JL & Tech.* 24 (2010), 171.

[59] Jatinder Singh, Christopher Millard, Chris Reed, Jennifer Cobbe, and Jon Crowcroft. 2018. Accountability in the IoT: Systems, Law, and Ways Forward. *Computer* 51, 7 (2018), 54–65. https://doi.org/10.1109/MC.2018.3011052

[60] Nili Steinfeld. 2016. 'I agree to the terms and conditions': (How) do users read privacy policies online? An eye-tracking experiment. *Computers in human behavior* 55 (2016), 992–1000.

[61] Heidi Tworek and Paddy Leerssen. 2019. *An Analysis of Germanys NetzDG Law.* Technical Report. Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression, Amsterdam, Netherlands. 11 pages.

[62] Jennifer Urban and Laura Quilter. 2006. Efficient Process or Chilling Effects-Takedown Notices Under Section 512 of the Digital Millennium Copyright Act. *Santa Clara Computer & High Tech Law Journal* 22, 1 (2006), 621–693.

[63] Jennifer M. Urban, Joe Karaganis, and Brianna Schofield. 2017. *Notice and takedown in everyday practice.* Technical Report. UC Berkeley, Berkeley, CA, USA.

[64] Ben Wagner. 2014. The Politics of Internet Filtering: The United Kingdom and Germany in a Comparative Perspective. *Politics* 34, 1 (Feb. 2014), 58–71. https://doi.org/10.1111/1467-9256.12031

[65] Ben Wagner. 2016. Algorithmic regulation and the global default: Shifting norms in Internet technology. *Etikk Praksis Etikk i Praksis* 10, 1 (2016), 5–13. OCLC: 6158737611.

[66] Ben Wagner. 2016. *Global Free Expression: Governing the Boundaries of Internet Content.* Springer International Publishing, Cham, Switzerland.

[67] Stephen C. Webster. 2012. Low-wage Facebook contractor leaks secret censorship list | The Raw Story. https://www.rawstory.com/2012/02/low-wage-facebook-contractor-leaks-secret-censorship-list/

[68] Liz Woolery, Ryan Budish, and Kevin Bankston. 2016. *The Transparency Reporting Toolkit - Guide and Template.* Technical Report. Open Technology Institute and Harvard University Berkman Klein Center for Internet & Society, Washington D.C. 64 pages.

[69] Rebecca Zipursky. 2019. Nuts About NETZ: The Network Enforcement Act and Freedom of Expression. *Fordham International Law Journal* 42, 4 (2019), 1325. https://ir.lawnet.fordham.edu/ilj/vol42/iss4/7