Pilot Project

DECODING ANTISEMITISM:

An AI-driven Study on Hate Speech & Imagery Online

First Discourse Report

Center for Research on Antisemitism (ZfA),
Technische Universität Berlin
King's College London (KCL)

Dr Matthias J. Becker (ZfA)
Hagen Troschke (ZfA)
Dr Daniel Allington (KCL)

ZfA Zentrum für Antisemitismusforschung · Technische Universität Berlin · KING'S College LONDON · ALFRED LANDECKER FOUNDATION

Table of Contents

1. Introduction

The interdisciplinary project "Decoding Antisemitism" facilitates the investigation of antisemitism on the internet. From the summer of 2020, this three-year pilot project will be based at the Center for Research on Antisemitism (ZfA) at TU Berlin and carried out in cooperation with King's College London (KCL). The object of investigation is antisemitism in the comments sections of media and social media platforms of the political mainstream[1] in Germany, France and Britain. In several research teams, a mixed methods approach analyses the data qualitatively, quantitatively and aided by AI.[2] The research project is supported by an international and transdisciplinary advisory board.[3]

In view of its multi-stage research design, but also its multilingual orientation, the project stands out from previous studies on antisemitism on the internet.[4] The multi-stage design enables the consideration and comprehensive examination of the different verbal and visual forms of antisemitism; the multilingual orientation enables contrastive representations of discursive trends in the three European countries.

By investigating comments sections in politically moderate milieus, we are opening up a field that has hitherto been neglected. In particular, we are able to represent in great detail the characteristics of this object of investigation in relation to antisemitism on the basis of extensive qualitative analyses. This thus provides an overview of the manifestations and extent of antisemitism in these European web communities which has been lacking up to now. The knowledge that can be derived from this survey also gives rise to possibilities for diverse forms of intervention.

The great significance of this research project is underlined by observations that show that the interactive web currently represents the most important place for gathering information and exchanging views,[5] that hate speech can lead to physical violence[6] and that attitudes expressed prominently and frequently on the Internet may

---

[1] In view of radicalisation trends in Europe and on the internet it seems obvious to focus on radicalised milieus, but we see a much greater danger in the normalisation of derogation and exclusion in politically moderate, socially established environments, whose transformations must be scientifically investigated.

[2] See https://www.alfredlandecker.org/en/article/decoding-antisemitism and https://www.tu-berlin.de/fakultaet_i/zentrum_fuer_antisemitismusforschung/menue/forschung/decoding_antisemitism_an_ai_driven_study_on_hate_speech_and_imagery_online.

[3] Prof Johannes Angermuller, Discourse, Languages and Applied Linguistics, The Open University, Milton Keynes; Dr Ildikó Barna, Department of Social Research Methodology, Eötvös Loránd University, Budapest, Prof Michael Butter, American Literary and Cultural History, University of Tübingen; Prof Manuela Consonni, Vidal Sassoon International Center for the Study of Antisemitism, Hebrew University; Prof Niva Elkin-Koren, Faculty of Law, Tel Aviv University; Prof Martin Emmer, Institute for Media and Communication Studies, FU Berlin; Prof David Feldman, Pears Institute for the Study of Antisemitism, University of London; Dr Joel Finkelstein, Network Contagion Research Institute, Princeton; Dr Claudia Globisch, Institute for Employment Research, Nuremberg; Shlomi Hod, HIIG's AI & Society Lab, Berlin; Prof Günther Jikeli, Institute for the Study of Contemporary Antisemitism, Indiana University Bloomington; Dr Lesley Klaff, Department of Law & Criminology, Sheffield Hallam University; Prof Jörg Meibauer, German Language and Linguistics, University of Mainz; Prof Eli Salzberger, Minerva Center for the Study of the Rule of Law under Extreme Conditions, University of Haifa; Dr Charles A. Small, ISGAP; Dr Abe Sweiry, Home Office UK; Prof Gabriel Weimann, Department of Communication, University of Haifa; Dr Mark Weitzman, Simon Wiesenthal Center; Dr Juliane Wetzel, ZfA, TU Berlin; Prof Matthew L. Williams, Criminology, HateLab, Cardiff University; German Research Center for Artificial Intelligence (DFKI), Berlin.

[4] On the necessity of a multi-level methodical approach, see Becker 2019 and 2020.

[5] On Germany see https://de.statista.com/themen/3207/internetnutzung-durch-kinder-und-jugendliche or https://de.statista.com/statistik/daten/studie/13070/umfrage/entwicklung-der-internetnutzung-in-deutschland-seit-2001 (last accessed: 14 January 2021).

[6] See among others: Ducol et al. 2016, Müller/Schwarz 2020, Williams et al. 2020, https://www.cardiff.ac.uk/news/view/1547823-new-technology-to-monitor-anti-polish-hate-online (last accessed: 14 January 2021).
On media reports see https://www.theguardian.com/us-news/2018/oct/27/pittsburgh-shooting-suspect-antisemitism, https://www.theguardian.com/technology/2018/oct/29/gab-forced-offline-following-anti-semitic-posts-by-pittsburgh-shooter,

views, that hate speech can lead to physical violence and that attitudes expressed prominently and frequently on the internet may have a societal dimension. In view of the tremendous spread and everyday nature of online media, the rigid separation between *online* and *offline* cannot be maintained.[7]

Project goals

The *Decoding Antisemitism* project is a multinational and interdisciplinary research project that aims to examine the explosive cross-societal problem of antisemitism in its manifestations, distribution and scope across national borders – and to make the results available beyond scientific interests as tools for intervention and prevention. In the course of continuous research of numerous web discourses, the project will catalogue antisemitic hate speech with a focus on implicit, elaborated and innovative forms and process the data obtained using different analysis methods. The main problem of previous studies is the insufficient consideration given in particular to the implicit as this limits the reliability of statements made about the presence and distribution of antisemitism in certain discourse spaces.

The research project is aiming to expand contemporary antisemitism research and will emphasise the need to bring together humanities on the one hand and AI and web-related disciplines on the other.[8] The possibilities that emerge from looking at online discourses using a number of different methodological approaches will send momentum towards the humanities and social sciences as well as towards research areas that are already concerned with the automated recording of hate speech, but – as we will later show – neglect central aspects that thus prevent it from being analysed in full.[9]

In addition to the research output, results on current trends in the three countries will also be compiled in half-yearly **discourse reports**.

The observable increase in antisemitic hate speech online (see e.g. ADL 2018 and 2019, CST 2019) and its impact on those directly affected as well as the political culture and character of society make research on antisemitic attributions and trends imperative. Therefore, we will use the discourse reports to make the results of our research studies accessible and usable for practical discussion of antisemitism beyond the research base for current antisemitism to support the development

---

https://www.theguardian.com/world/2019/oct/09/germany-shooting-synagogue-halle-livestreamed, https://www.washingtonpost.com/technology/2019/03/15/facebook-youtube-twitter-amplified-video-christchurch-mosque-shooting (last accessed: 14 January 2021).

[7] On the USA, see Nagle 2017; on Germany, see Stegemann/Musyal 2020. See also media socialisation research such as Kneuer/Salzborn 2016, Hoffmann/Krotz/Reißmann 2017 and Rieger 2019.

[8] An initial article that deals qualitatively with the antisemitic attributions against George Soros, which we briefly present in Section 5.1, is currently in preparation and will appear this year in the Journal of Contemporary Antisemitism. These results, among others, will be utilised within the framework of the aforementioned interdisciplinary merging of the project.

[9] Previous studies have generally opted for a qualitative (Becker 2018/2021, Jikeli et al. 2019, Schwarz-Friesel 2019) or quantitative (ADL 2019, Barna/Knap 2019, CST 2019, Zannettou et al. 2020) approach and their object of investigation is limited to one or a few media or platforms in order to research antisemitism on the internet. In particular, the quantitative studies are characterised by the fact that they are only able to identify novel and above all implicit forms of expression (at the word and sentence level) to a limited extent. This leads to problems of distorted results or incomplete recording of antisemitic statements – especially in relation to linguistically moderate mainstream discourses. Joint qualitative and quantitative analysis can be found in ADL 2018, Allington 2018, Woolley/Joseff 2019, Allington/Joshi 2020, Ozalp et al. 2020.

of strategies and decisions (e.g. to review existing content moderation concepts or design learning formats about antisemitism).

In addition to mapping current processes in the reports, another output of the project will be an open source tool. This is being developed based on the precise categorisation of antisemitic statements and will be made available to the moderators of selected platforms and news websites. In future years, an **internet institute** representing the *Decoding Antisemitism* project is expected to be established. This will provide an institutional interface between research and the areas mentioned above and stabilise the transfer of knowledge, especially as the examination of antisemitism will be expanded to other European countries after the pilot phase.

Current status and outlook

After the project was given the official go-ahead in August/September 2020, a two-person team oversaw the organisational installation of the project and carried out initial data surveys and qualitative corpus analyses over two months.

Once all the country teams have been established in the coming months, the originally planned format of research work will begin, i.e. the continuous implementation of multimodal, contrastive analyses of antisemitic statements and trends in the three European countries. Important discourse triggers in the three countries will be determined and, in a later project phase, the analyses that up to now have been linguistically oriented will be supplemented by investigations of visual patterns.[10] Based on the qualitative categorisation of extensive amounts of data, the transfer to AI-based and

quantitative analysis steps described in section 6 can also take place, the majority of which will be carried out at KCL.

Accordingly, the first contrastive-oriented report with results obtained from qualitative, AI-based and quantitative analyses will be submitted in summer 2021.

2. Data collection

Our data is based on user comments, which we collect from the comments sections of the websites and *Facebook* pages of mainstream media (in particular newspapers) from Germany, Britain and France and store in anonymised or pseudonymised form. With these we want to determine the nature of the existence of antisemitism in discourse spaces of the supposedly moderate political milieu.

In addition, reference corpora are created that are structured according to the same criteria as the main corpus (initial stimulus political topics), but in which the preceding articles do not contain the search terms for the main corpus. Contrastive comparisons between main corpus and reference corpora make it possible to determine the presence of antisemitism depending on certain discourse triggers.

After completion of the project, both the raw and analysed data will be archived for documentation purposes, controllability and reuse (for a limited group of users with proven research interest through authentication).

---

[10] In the web milieus with which the research project is primarily concerned, antisemitism mainly appears in verbal (and not visual) form. A focus on linguistic and discourse-analytical investigations is therefore logical, especially as the degree of implicitness in language use plays a major role.

3. Study design

This investigation follows a methodical three-step process: first, the teams examine current antisemitic patterns of language use within the framework of Mayring's **qualitative content analyses** (2015). The categories are developed inductively (Meibauer 2008) both in relation to antisemitic concepts (including stereotypes in particular; Schoeps/Schlör 1996, Julius 2010) as well as the linguistic and visual phenomena employed by users. In addition to this categorisation, catalogues are also developed that show the combination of conceptual units and linguistic-image patterns in each language community.

The second step is to use the results of these qualitative analyses for **machine learning** at KCL. A precise and context-sensitive categorisation of the web comments provided in the qualitative step aims to use the categorised data as training data in the development of an algorithm. Its target is to carry out the task of distinguishing between (clear) antisemitic and non-antisemitic comments with increasing accuracy over the project period. To this end, continuously updated data, which can also reflect changes in the antisemitic repertoire, are fed into the machine learning process during the course of the project. Automation allows much larger data sets to be examined. It is possible to record which stereotypes are communicated via which patterns in which web environment. Due to the complexity of language use and the associated challenges for an algorithm to classify texts correctly, this step will be continuously supervised and include numerous test runs. In this context, different variations of machine learning models will also be tested.

In a final step, **quantitative analyses** (focusing on frequency, collocation and vector analyses) are carried out. These are based on the results of the previous analysis steps. The analysis of the data is carried out by three country teams[11] at the ZfA and one researcher at KCL.

The results of steps 2 and 3 ultimately serve to bolster quantitative analyses using updated catalogues that collect current antisemitic communication patterns.

Why did we choose this exact methodical arrangement? It would also be possible to initiate the investigation using quantitative analyses to comb through large data sets and then, with these findings, continue on to detailed qualitative analyses. The reason for our decision to start with the latter is due to the perennial observation that antisemitic (including group-related) derogation and/or exclusion generally take place implicitly in segments of the political mainstream (see among others Schwarz-Friesel/Reinharz 2017, Becker 2018/2021).[12] Verbalisation variants sometimes differ to a great extent from expected patterns (such as those that appear in far-right contexts, for example). This means that when searching for relevant words (swear words, explicit death wishes or the open reproduction of antisemitic stereotypes), a large proportion of the posts in which antisemitism is communicated cannot be taken into account. Even vector analyses (in which accumulations and combinations of words are investigated within a medium and reference period, i.e. researchers are guided by corpus-specific frequency distributions and not by predetermined, deductive categories, e.g. ADL 2019 and Zannettou et al.

---

[11] One postdoc researcher and two PhD students will deal with current phenomena and trends for each language community. In addition, the teams cooperate for contrastive studies of the country-specific discourses.
[12] There are three possible causes for the communication of antisemitic attributions in an implicit, coded way, i.e. via abbreviations, puns, allusions, etc. which in turn presuppose context-dependent knowledge: this way of conveying information can increase its

2020) cannot capture the numerous comments in which antisemitic stereotypes are reproduced without recourse to relevant-specific word selections (see section 5). In other words, qualitative analyses impressively demonstrate that comments sections where a predefined search for deductive categories resulted in few or no hits can nevertheless contain large numbers of antisemitic statements; they also illustrate that the constitution of antisemitic attributions can be semantically so open that conspicuous word accumulations – or even relevant terms – can be completely absent.

This shows that the interlinking of the three methodological steps proposed here will enable us to analyse antisemitism on a large-scale online, while also maintaining the accuracy of categorisations. The Decoding Antisemitism project will ultimately reveal reliable results on the shape and frequency of antisemitism on the internet of the three language communities.

4. Code system

In the MAXQDA analysis tool, a categorisation or code system is created to map the (content-conceptual, lexical-semantic and pragmatic) diversity of the object of investigation as well as its respective scope.

Antisemitic stereotypes represent mental units that have been handed down through language and images, sometimes over centuries. Accordingly, classical and modern stereotypes (such as EVIL, GREED, DECEIT, LIE, POWER[13]etc.) as well as conspiracy theories, forms of Nazi analogy and other categories of antisemitism have their place in the code system. Its conceptual repertoire is amply defined in antisemitism research. In the extrapola-

tion of country- and milieu-specific debates however, we follow an inductive category formation in order to be able to include any novel attributions (as well as their distribution in the individual web milieus).

Another part of the code system refers to the level of language, i.e. the question of which word choice, sentence structure, speech acts, symbols, etc. are used to reproduce the respective stereotype. As mentioned above, when recording antisemitic attributions, we aim to make multiple codings, i.e. for each coding of a comment to take into account ideally (or in the case of linguistic-visual peculiarities of a stereotype) the linguistic-visual level in addition to the conceptual level.

In concrete terms, this means that the example statement Will they ever have enough? – in which it is apparent from the co- and context that 'they' refers to Jews – is coded several times, namely: Greed (= antisemitic stereotype; content-conceptual level) and rhetorical question (= indirect question; pragmatic level). In order to guarantee the categories and interpretations of the passages, a consensual validation is carried out (Bortz/Döring 2006) and intercoder reliability is calculated. The extrapolation of the meaning of implicit statements results from the combination of knowledge regarding conventionalised forms of language as well as contextual, cultural and world knowledge.

Finally – in addition to concept and linguistic pattern – the levels of the semiotic and visual units should be mentioned as they are also decisive in constituting communicative sense or are given extra importance, especially on social media platforms. They include emoticons, GIFs, memes and other kinds of images, as well as typographic characters.

---

effectiveness for the recipient (since, in this way, writers allow readers to participate in "secret knowledge"); it follows aspects of language economy (preferring the implicit forms for the sake of efficiency); it serves to protect the reputation or the self-image of the writer (if this is seen as a necessity in certain milieus) and it avoids sanctions from moderators.

[13] Since stereotypes are phenomena that exist on the conceptual, i.e. mental, level and can be reproduced using language, stereotypes are given in small caps on the following pages in accordance with the conventions of cognitive linguistics.

5. First corpus analyses

In the first phase of the project, we looked at the comments sections of German and British mainstream media and their profiles on the social media platform Facebook and selected three relatively current discourse triggers for the analyses: debates about George Soros in the first half of 2020, the publication of the EHRC report on Labour antisemitism in October 2020 and the expulsion of Jeremy Corbyn from his party and debates about coronavirus in autumn 2020.

In our analysis of web debates on Soros, we followed the qualitative content analysis method discussed in sections 3 and 4. This was also used in the investigation of online debates relating to the EHRC report though here there were additional keyword searches. Where there were hits, the corresponding text segments were analysed in detail. For the third discourse trigger, we looked at a large data set and, in view of the volume of data, carried out qualitative samples then extrapolated the entire data using the quantitative approach of the keyword search and continued the investigation with detailed analyses. On the basis of empirical material, the results of these three corpus analyses demonstrate the importance and suitability of the study design for extrapolating the object of investigation.

5.1 British media coverage of George Soros

The well-known Jewish billionaire, investor and philanthropist George Soros has been the focus of antisemitic attributions for several years (Barna/Knap 2019, Woolley/Joseff 2019). Moreover, reports emphasised the presence (partly relating to Soros) of antisemitic statements in the context of coronavirus (ADL 2020a and b, CST 2020).

In our first qualitative corpus analysis, we aimed to shed light on this link with a view to the coverage of Soros in the UK.[14] How do web users use coverage of Soros in the British mainstream media as an opportunity to spread antisemitic attributions? In the corpus, which consists of comments sections from the *Daily Mail*, *The Guardian* and *The Independent*, we were only able to find a few clear claims in which Soros was blamed for the pandemic or the emergence of coronavirus (see also section 5.3). It seems to be primarily a phenomenon of radicalised milieus which were the focus of the above-mentioned studies. Nevertheless, the comments sections examined here contained antisemitic attributions and cross-connections between them, which we will briefly present below.

A variety of classically **antisemitic stereotypes** came to light in the examined comments sections. For example, users repeatedly attest in a direct manner that Soros represents evil. Users also explicate their view that Soros is close to the devil, use animal metaphors (known from the history of antisemitism) and produce, among other things, a contrastive juxtaposition between Christians and Jews to the detriment of the latter. These kinds of **constructions of difference** are also produced within other in-groups. Thus Soros is alleged to be the antagonist of the individual but also of entire nations and even the world. Here we come full circle to other stereotypes: Soros is alleged to be harming humanity, his behaviour is also characterised by greed, deceit and lies and he has an extraordinary and uncontrolled power. He is said to use this power to influence public opinion or political events to his own ends and/or to push through his own hidden interests. This are exactly the allegations which have historically been made against Jews collectively, as in *The Protocols of*

---

[14] The results of this corpus analysis will be published in detail in the next issue of the peer reviewed *Journal of Contemporary Antisemitism* (JCA). Some results are already presented in this report, even though they are normally only communicated to the outside world in a scientifically validated manner as part of a peer review process.

the Elders of Zion. The historically well known accusation of DISINTEGRATION appears in the corpus, as does the claim that various **conspiracies** originated from him. The attribution that he does not follow any morality takes on an extreme dimension in the comments when users insinuate that as a child he **collaborated with the Nazis in the Holocaust** and sent other Jews to their deaths.

The reproduction of stereotypes and the attribution of other negative characteristics are often **combined** or linked through arguments. For example, influencing public opinion (and/or political events) is said to serve his aim of disintegrating society and its values. **Death wishes** against Soros, which ultimately represent the most radical positioning – elimination antisemitism – are formulated in many ways, sometimes as calls for action.

Linguistic observation shows that, despite the high presence of verbal aggressiveness and open rejection of Soros, users use a variety of coded forms to communicate their antisemitic attributions. Although a considerable proportion of stereotypes such as EVIL, DEVIL and MENDACITY were coded using relevant words (as well as sometimes highly derogatory swearwords and insults), all attributions show a wide range of implicit, elaborated and innovative forms of expression. Starting with semiotic markers (e.g. icons and emoticons as in "🤷oro🤦🤔"), commenters use abbreviations ("NWO"), wordplays and allusions ("Deep state")[15] as well as metaphors (as in "[his eye bags] are money bags") and contrastive juxtapositions as mentioned above ("Soros the devils advocate is worried, must be good then"). In terms of the speech act of the curse or death wish, there are both explicit forms ("just die already") and coded forms ("Someone needs to give Soros a 'shower'"

as an allusion to the gas chambers). Users also use indirect speech acts (such as rhetorical questions and irony): "Is he trying to interfere in the 2020 election?" or "Like, we can believe ANYTHING from George Soros?" which reproduce antisemitic attributions even more subtly. Extrapolating the meaning here in many cases results solely from the combination of knowledge of conventionalised forms of language as well as contextual, cultural and world knowledge.

In the case of these diverse linguistic patterns, the use of relevant words (for the production of antisemitic attributions) is often completely absent. This confirms once again that in a purely quantitative investigation of antisemitism, a large number of those statements that contain implicit hate speech cannot necessarily be collated. In order to be able to fully identify the numerous varieties using quantitative analyses, a larger set of key terms would be needed. The small number of relevant terms would make a quantitative search for these passages extremely complex. Depending on the method, this would have to include many words and combinations of them – and even then not all antisemitic statements could be captured. This is evident from the indirect speech acts mentioned here in which the meaning is constituted by several elements of expression and their interpretative content (and not concentrated at the explicated word level).

Preliminary searches for relevant antisemitic vocabulary in our corpus suggested that antisemitism is negligible in these comments sections of the three British media. A **quantitative classification** carried out after the qualitative evaluation of overall 1,244 comments shows the actual extent: 15.2% of the comments were antisemitic. This val-

---

[15] This topos gained popularity through the widespread use of the meta-conspiracy theory *QAnon* which is interwoven with various other conspiracy theories. According to this, members of a globally controlling elite committed crimes against children and Trump is said to want to stop them and curb their power.

ue was more or less the same for the left-wing media The Guardian and The Independent (15.0%) as well as the right-wing Daily Mail (15.4%).

5.2 EHRC report on antisemitism in the Labour Party

Another discourse trigger for generating antisemitism that we considered was the publication of the EHRC (*Equality and Human Rights Commission*) report on antisemitism in the British Labour Party on 29 October 2020 and the subsequent expulsion of former chairman Jeremy Corbyn from the party.

We analysed *Facebook* threads that could be accessed on the profiles of Corbyn himself as well as those of *The Guardian* and *The Independent*. In contrast with the procedure presented in section 5.1, this time it is not a consistent detailed analysis of the developments in a series of threads but rather an exploratory investigation in which the first 100 user comments per thread were examined qualitatively followed by – in order to identify further antisemitic comments – a search of the entire corpus using relevant topic area terms. Where we found hits, the cotext and the dialogic processes were then examined in detail. In the total of nine *Facebook* threads, 1,272 comments were qualitatively examined,[16] 17.2% of which contained antisemitic statements.

A common feature of all *Facebook* threads was the strong presence of the stereotype POLITICAL (and partly also MEDIA) INFLUENCE – be it from Jews in Britain or from Israel or a so-called Zionist *lobby*. Users speak of a general "pressure" on politics emanating from Jews[17] and a "victim mentality" (which is said to ultimately have negative

consequences, especially for Jews themselves). This combines the stereotypes of POWER, INSTRUMENTALISATION OF and, consequently, BLAME FOR ANTISEMITISM: "I actually worry this course of action will increase mistrust of the Jewish members, as it becomes increasingly obvious they have exerted pressure and used a victim mentality to oust the best hope labour had in a generation." Alongside the accusation that Jews had set themselves the goal of overthrowing Corbyn and damaging the Labour Party, many comments suggest that they also had the ability to achieve this goal. In general however, it is claimed that such INFLUENCE originates directly from Israel: "the israhell lobby flexes its nasty influence again!" "The israeli lobby has captured the media" – or more subtly: "Labour Party is totally controlled by a foreign State, same with the Tories."

The threads generally feature numerous references to the Middle East conflict, in which users – in addition to legitimate criticism of the Israeli government – communicate antisemitic stereotypes and demonising analogies.[18]

The claim that Israel is behind Corbyn's exclusion extends as far as users employing Nazi comparisons with regard to Israel and accusing the country or its inhabitants of mendacity: "So anti semitism started and ended with Corbyn. [...] Stop this nazi style Israeli lying."

The concept of INTERFERENCE by Israel is also communicated indirectly – and can be extrapolated through the use of emotion words and the co-text linking Israel to Corbyn's exclusion: "This is what happens when you love your country more than Israel." Using the rhetorical question "why would anyone Palestinian or any of us fight for the end

---

[16] The results of this corpus analysis will be presented and explained in detail in a peer reviewed journal during the course of 2021. The same remark as in footnote 14 applies here.
[17] Notably Luciana Berger but also other former Jewish members of the Labour Party are mentioned in the comments sections.
[18] This happens even though the EHRC report primarily problematises discrimination against Jewish people within the Labour Party (and not events in the Middle East). The report derives from a long-standing need for clarification on antisemitism in the Labour Party, which can be seen at the political, social and media level (see Rich 2016, Allington 2018, CST 2019).

of Izrahells apartheid and systematic genocide go near a labour party that has been literally hijacked by Zionists??" – with the apparent focus on demonising Israel by means of puns (Izrahell) and injustice scenarios (genocide, apartheid) – the insinuation of a Zionist influence on British politics is presupposed quite incidentally.

The suspicion of INTERFERENCE also proceeds through content-related circumlocutions, with users explicitly claiming that Israel's interference involved right-wing segments of the Labour Party. Keir Starmer in particular, as the new Labour leader, is repeatedly perspectivised as a stooge, a "puppet" and "servant" of Israel (and of the Tories). The image of the Jewish puppet master controlling politicians is familiar from the history of antisemitic propaganda.

The perspectivisation of current events as (internal or external) INTERFERENCE raises the question of how to deal with the thematic core of the report: antisemitism in the party. In many comments, users demonstrate that they do not take references to this seriously and relativise or even deny Jew-hatred in their own ranks. According to users, underlying the subject of antisemitism is the function of an INSTRUMENTALISATION (and thus the CONTROLLING OF POLITICS AND THE PUBLIC) ("you're weaponising anti Semitism"), which also reproduces the stereotype DECEIT. Users believe that the "Israel lobby" ought to be given the BLAME FOR ANTISEMITISM: "Nothing caused more antisemitism in the Labour Party than an aggressive Israeli lobby angry with the election of a pro Palestine socialist labour leader , and acting in collusion with the party's right wing and a compliant media to push fake antisemitism stories in order to get rid of Corbyn." Antisemitic incidents in the party's past are once again denied here and problematising them is viewed as a political tool.

In producing the attributions, users' apparently strict separation of *Jewish, Zionist* and *Israeli* is repeatedly breached: "I left the party, when the

Jewish Board of Deputies issued their '10 commandments'; it was obvious to me then that the Zionist Lobby were out for the blood of Socialism." Imputed injustice is linked to Jewish tradition: when one user asks, "what's the difference between a Zionist and a Jew?" another replies, "You don't have to be a Jew to be zionist and thank God not all Jews are Zionists [...] Clearly you're a zionist... Therefore you are the worst kind of racist.. the Talmud makes that clear."

Conspiracy theories also play a role when looking at the EHRC report. In the following example, insinuations of POWER and INFLUENCE in connection with dichotomous patterns of world interpretation (already identified in the Soros corpus) are brought together and a submission to a "foreign entity" is volubly implied using allusions and metaphors:

> "if Jeremy Corbyn can't do a thing about it then there is absolutely NO hope left. A misnomer asking decent socialists to stay in a party that is no longer socialist but **ruled by Zionists**. A new socialist party for people with REAL socialist values is needed for the people not a Tory light party who **pander to a foreign entity**. [...] People joined in their thousands because of a gentle man who only had the good of the people in his heart but lacked the backbone to get rid of the **'wolves in sheep's clothing'** Who on earth would want the likes of **Berger, Ellman taking bungs from the Israel lobby, Austin** etc back, beggars. Poison Hodge and Riley orgasming over what's happened, makes me sick to my stomach. That ain't no party for the people, they're all **careerists kowtowing to the Likud party**. Disgraceful!!!!" [authors' emphases]

In addition to the allegory of wolves and sheep and the reference to Jewish members who left Labour because of antisemitic incidents, the contrastive juxtaposition of *people* reflects another construction of difference. In the above example, Israel and British Jews are conceptualised as relat-

ed actors. This idea also appears in the following dialogue: A: "Israel owns both our main political parties, so now even if the conservatives lose an election, THEY STILL WIN! 🤣 🤣 🤣," B: "it's the Jews!" Irrespective of this separation, however, here there is the fundamental antisemitic attribution that alleges Jews control the parties.

According to users, the imputation of a conspiracy does not relate solely to the British political landscape (as in "UK becoming a vassal state of Israel. Jewish Zionist lobby is most powerful organisation in the UK which has power to destroy parties, bring new Home Secretaries and practically do what they wish") but also to other states: "Israe-hell have the worlds leaders in their pockets." In the following, the idea of a Jewish-Zionist conspiracy is communicated in a semantically open way – "The powers-that-hope-to-be tried the same to Bernie Sanders although he is Jewish" – however the concessive clause and the statement's cotext enable a conclusion to be drawn about Jews as actors (also in the USA). Users also use animal metaphors that communicate the topos of the world conspiracy in traditional metaphors – and which are comparable to the *Stürmer's* methods: "aligned with Zionist Israel whose tentacles stretch here and everywhere."

The topoi highlighted here and which are central in the EHRC corpus show how antisemitic stereotypes and the idea of a (world) conspiracy are produced linguistically and can be linked to each other. Like the statements presented in section 5.1, the comments exhibit a variety of language usage patterns. By extending attributions to encompass *all* Jews, users repeatedly overturn the oft-mentioned justification in the debates about Labour antisemitism that people refer solely to Israel's dealings with Palestinians. They use the debates surrounding the EHRC report as an op-

portunity to imply a secret plot that allegedly affects not only Britain but the whole world and as a result (aided and abetted by their own politicians) establishes a taboo on criticism regarding Jews and Israel. If ones takes into account accompanying emotions, it can be seen that in debates on Soros a higher proportion of derogatory ridicule is observable – in the corpus discussed here, however, many statements are backed up by feelings of impotence and anger.

5.3 Web debates on the Covid-19 pandemic

On the basis of the observation that in coronavirus-related discourse conspiracy theories arise about the nature of the virus and its spread, as well as other forms of antisemitism,[19] we looked at the extent to which these theories are also found in politically moderate discourse spaces. To this end, based on the media we selected, we compiled a corpus from the comments sections under those articles that contained the terms "Corona" or "Covid" and under which users had written comments. For the investigation, we chose one of the most prominent time periods since the beginning of the coronavirus-related events in Germany (26 October–2 November 2020). This is an exploratory analysis of a German-speaking corpus and the aim is to extend this in a next step to include further investigation periods and contrastive corpora from the other language areas.

Under the 65 articles that met the selection criteria, 17,476 comments were published. Of these, 2,000 comments were qualitatively analysed on their content-conceptual subject matter relating to antisemitism and conspiracy theories and, if necessary, on their linguistic expression. Thematically, the comments intensively discussed, among other things, the measures to contain the pandemic and this resulted in decision-makers

---

[19] ADL 2020c, Bundesamt für Verfassungsschutz 2020, Roose 2020, Shahrezaye et al. 2021.

being blamed for (assumed) mistakes.

During the analysis, it became apparent that, contrary to expectations, there was virtually no antisemitic or conspiracy-theoretical content in the individual threads. Since it was quite possible that such content was rare and therefore hardly to be found in the coded extract, a quantitative analysis of the entire corpus was subsequently carried out to see whether certain key terms appeared that can be regarded specifically as an indicator of (coronavirus-related) conspiracy theories and/or antisemitism (or arose from the data as potential terms for alleged conspiracies). This second method was used in order to extrapolate more extensively from the corpus. It showed that these key terms were mainly found either in a neutral context or in counter-speech. In counter-speech, they were used to contradict the positions suspected by their addressees (almost no one plainly confessed to conspiracy thinking).

Of the analysed comments, we found that there were six that propagated a structurally antisemitic conspiracy theory. These assumed that an unnamed elite is using the pandemic to implement plans for world domination or to introduce a new order: "In fact, the pandemic is being used to implement an agenda that would be difficult to enforce without it. [...] The pace at which [politics] is proceeding shows how far the elites' plan to reorder the world in their favour has already come." The idea that a particular group allegedly wants to implement a global agenda and thus be able to shape the world according to its interests corresponds, with its combination of attributed goals, actions and the necessary power, to a central narrative structure that is traditional in antisemitism. From this analogous narrative structure it emerges that the semantic range of possible interpretations for the agent *elites* is limited and that it is to be regarded as an equivalent for *Jews*. The same is

true of the following statement with regard to a *clique* that is said to hold the German Federal Government in its power. With *puppets*, however, there is a further reference to an antisemitic meaning as this is a relevant metaphor that implicates full external control: "The puppet regime in Berlin is merely a lowly stooge of the clique."

There were no other kinds of antisemitic statements in the comments. In 16 other comments, users attached conspiracy theories that are not necessarily antisemitic. They referred to an all-powerful regime, governments, elites, Bill Gates, among others. However, they were left under-specified in attributions regarding actions, objectives or motivations. Furthermore, these comments sometimes contained terms that suggest a conspiracy-theoretical interpretation since these are common in this milieu: "totally right. This has long been seen in the increasing digitalisation, 5G, artificial intelligence... But we are too weak against these elites... I don't know how we could stop this process..." As above, *elites* are being discussed here. The implication is that certain people are (over)powerful and responsible for certain developments – but without a clear narrative that would place the term in an antisemitic context. Even more subtle is the idea of a conspiracy regarding the precautions against Covid-19: "I am not worried about a virus, but rather the disproportionate measures. I'm not falling for the test 'pandemic'. nor for the propaganda that is nowadays called a 'fact check'." By criticising the measures here, the effects of Covid-19 are depicted as harmless. The idea of a test '*pandemic*' is built on this, implicating that the official handling of the disease is instrumentalising it and using it as a pretext to enforce a trial run and the first phase of a transformation of society (*measures*). This is said to occur by means of support through the targeted dissemination of misinformation (*propaganda*).

There is thus a reference to concrete actions – but actors or targets are not mentioned. Fact-check[ing] is thereby, in an inversion of reality, issued as propaganda, while in the logical, still implicit reversal of this argument, the refuted misinformation, which essentially denies the importance of the disease and serves conspiracy theories, is raised to the level of truth.

Given the well-known links between conspiracy theories and discourses rejecting pandemic response measures, these findings exhibit an unexpectedly low presence in the discourse spaces studied. In principle, this is a relatively gratifying finding. However, it does not have to be solely due to the fact that such ideas are communicated to a small extent. It is possible that the moderators of the media are very reliable at removing corresponding comments. Another interpretation would be that, at least in mainstream discourses, the discussion of coronavirus measures does not really act as a trigger for antisemitism. Furthermore, it would be conceivable that explanations of the pandemic using clearly recognisable conspiracy theories or even references to *powerful Jewish elites* are avoided since they are considered too absurd in the discourse spaces under consideration and are more likely to be rejected by other users. It is noticeable that users sometimes comment with hints or fragments of conspiracy theories (e.g. on the subject of vaccination using statements with misinformation about the effectiveness of vaccinations per se or that of tests) but without demonstrably allowing themselves to be pinned down to the associated conspiracy theories. One can assume that they want to disguise corresponding positions – again in order to not meet with rejection or to avoid sanctions or to convey their ideas in small doses covertly and unnoticed. This communicative behaviour would be in contrast to the expression

of Israel-related antisemitism which meets with more approval and is communicated more openly under the pretext of criticism, or the media reports on Soros discussed in section 5.1 in which users can position themselves, among other things, as anti-capitalist.

This reflection does not however sound the all-clear. Antisemitism was and still is a phenomenon that manifests itself especially at times of crises (such as the current pandemic) – in radicalised milieus in many forms and probably high frequency. Trends in online coronavirus-related debates will continue to be the subject of our research. One of the aspects that will be of interest here is the extent to which the antisemitic attributions of radicalised milieus – with the currently observable increase in dissatisfaction in society as a result of the continuous lockdown measures – enter mainstream debates.

6. Quantitative studies of antisemitism

6.1 Automated hate speech identification

'Machine learning' refers to the use of computer algorithms which improve their performance on specific tasks through exposure to structured data. Typically, the tasks in question are such that would previously have had to be done by humans, for example driving a wheeled vehicle or responding to spoken instructions. For this reason, machine learning is considered to be a branch of artificial intelligence, although 'intelligence' is a misleading term for this sort of automated decision-making: machine learning systems do not understand what they are doing; they simply attempt to replicate (to a greater or lesser degree of accuracy) the decisions that suitably trained humans would have made under the same circumstances. In many cases, computers are unable to perform as reliably as

humans would, given sufficient time. However, they typically decisions at a much faster rate, and can do so 24 hours per day without a break. This does not remove the need for human involvement, but it can extend the capabilities of human operators.

Historically, many internet companies have demonstrated little commitment to the removal of hateful content, on occasion even making the argument that it would be unethical to do so (see Daniels 2018 for discussion). However, in recent years, the tide has turned, with major platforms acknowledging their duty to moderate content responsibly, even if minor platforms such as Gab and Bitchute present their lack of moderation as a unique selling point. However, even where internet companies sincerely wish to remove hateful content, the scale of the challenge is enormous. *Facebook*, for example, reportedly employs around 15,000 content moderators (Thomas 2020). Yet it is clear that 15,000 individual human beings would be unable to moderate the hundreds of millions of *Facebook* posts that appear every day. Because of this, platforms rely on machine learning as a 'force multiplier' for their human content moderation teams, using this technological fix both to identify content that may need a decision made, and to identify content that is similar to content that has already been sanctioned. *YouTube*, for example, has explained its approach as follows:

> "We normally rely on a combination of people and technology to enforce our policies. Machine learning helps to detect potentially harmful content, and then sends it to human reviewers for assessment. Human review is not only necessary to train our machine learning systems, it also serves as a check, providing feedback that improves the accuracy of our systems over time. Each quarter, millions of videos that are first flagged by our automated systems are later evaluated by our human review team

and determined not to violate our policies." (YouTube 2020)

With regard to the example of child abuse imagery, *Facebook's* similar approach is explained as follows:

> "Automated systems are already responsible for removing 97.5% of those types of posts that appear on Facebook. But according to Facebook spokesperson Emily Cain, human reviewers are critical when it comes to 'banking' [such] material. That is, taking known images and logging them so that Facebook's AI systems can then go find and remove them. 'Without humans banking this content then our machines can't find it at scale,' Cain said." (Lapowsky 2020)

Content moderation systems that integrate human decision making and machine learning are thus already a part of the internet's day-to-day functioning. However, there are two specific concerns with regard to hateful content that do not arise with regard to other forms of banned content such as child abuse imagery. Firstly, it is often difficult to draw the line between hateful content and legitimate opinion, and there may be considerable difference of opinion with regard to the categories involved. And secondly, those who wish to disseminate hateful content have learnt to exploit this ambiguity by expressing their opinions in subtle and coded forms – as this report has already shown.

Although it is often possible to appeal individual content moderation decisions, it is hard to have confidence in the decisions which platforms make at scale, because they are not transparent about the exact forms of guidance and training that are provided to their human content moderators. This is a particular problem when we come to forms of hate that are not well understood within the mainstream of the counter-hate industry, such as antisemitism. Whilst a scholarly research project could never hope to compete with the software

engineering capabilities of a major internet platform, there is therefore scope for the academic development of hate speech identification system founded on the decision-making of trained experts relying on theoretical principles grounded in a transparently articulated understanding of this specific form of hate. The purpose of this system will not be to replace the sophisticated content moderation systems already in place at major internet platforms but (a) to provide an example from which the architects of such systems may learn about the specific problems associated with less widely understood forms of hate and (b) to educate policymakers and civil society about the potentials, limitations, and risks of automated content moderation.

The internet industry is already engaged in automated or semi-automated content moderation on a vast scale. Our job is to help industry leaders and policymakers so ensure that it does the best job possible. This includes not ignoring antisemitism, not simplistically focusing on 'forbidden words', not ignoring implicit hate speech, and not mis-categorising legitimate opinion as hate speech.

6.2 Use of qualitative coding as structured data

Structured data consists of data (such as texts and images) which have already been classified by human experts, providing examples of 'correct' decisions for machine learning systems to replicate. The activity of this project's expert teams working to identify antisemitic discourse in multiple European languages does not only serve as a contribution to knowledge in its own right: it also serves to generate structured data for machine learning. Following best practice, the dataset will be divided into three parts: one for training of

machine learning algorithms, one for testing them to identify the most highly performant, and one for evaluating the performance of the most successful. The evaluative stage will be particularly important, as it is our aim to explore the limits of algorithmic content moderation with regard to this particular kind of hate speech – not merely providing a percentage success rate for replication of categorisations made by human coders, but identifying the types of speech that most typically lead to errors of categorisation.

In generating such data and using them in this way, the project acts as a microcosm of the algorithmic-human hybrid content moderation systems used by major internet platforms. By focusing on one specific, rigorously theorised form of bad content – that is, antisemitic hate speech – and by conducting every stage of its work with full transparency, it will provide unique insights into the potential of such systems for to improve online discourse, also unveiling the problems that arise, especially with regard to false positives and to the difficulty of keeping up with forms of coded discourse that evolve in response to existing content moderation regimes. Both classical machine learning approaches (such as support vector machines) and deep learning approaches (i.e. involving neural networks) will be experimented with in the course of this research, and structured data will be made available to other researchers. All code will be open sourced, and evaluative analysis of its output will be published on an open access basis. Because the categorisation system used by the project's team of 'content moderators' is so sophisticated, featuring (for example) codes not only for hate speech but for speech about hate speech, we will be able to provide detailed information on exactly where the solution that is ultimately developed performs most and least effectively.

6.3 Experimental study of antisemitism and its perception

Simultaneously with the collection, coding, and analysis of naturally-occurring data, the project team is also collecting experimental data on antisemitism and how it is perceived. This is a separate study with the goal to establish the psychological correlates of antisemitism.

The first such experiment was carried out from 30-31 October 2020, with help from Louise Katz of the University of Derby. A balanced sample of 809 adults was recruited through the Prolific platform, with approximately equal representation of males and females, and approximately equal representation of adults aged under and over 25 (the sample was balanced in this way rather than for age representativeness of the population as a whole in order to control for recognised psychological differences between young adults and older adults). The purpose of this experiment was to clarify the relationship between antisemitism and a wide range of other traits, as well as to provide validation data for a new scale for the measurement of antisemitism. The findings, which shall be outlined in brief below, will inform the design of future experiments (in particular, by identifying sub-populations that may need to be focused on in more detail and psychological and other factors which may need to be controlled for) and will also be written up as a series of articles for peer review, as their inherent interest to the academic community is so great.

Findings will take some time to emerge, as the dataset is complex and requires careful analysis. Moreover, it is generally wise to avoid releasing potentially controversial findings before publication in an academic venue, as it is through peer review that research is certified. However, a number of findings can be reported.

Firstly, it has been possible to use the data to validate a new measure of antisemitism, the Generalised Antisemitism or GeAs scale. This was developed from two previous measures for which good data already existed. These were the established questionnaire instrument used in the Antisemitism Barometer survey which has been carried out annually by the British voluntary sector organisation, Campaign Against Antisemitism since 2015, and a new instrument developed from 2018-2019 in order to measure forms of antisemitism recognised in the IHRA Definition but previously ignored by policymakers (Allington/Hirsh 2019). Items from the two instruments were combined and standardised on the basis of existing data, and then tested using principles associated with Item Response Theory. It was found that a bifactor model – that is, a model which assumes that questionnaire items measure a single underlying factor in addition to two or more 'group factors' – provided the best fit for the data. This supports the view that 'classic' antisemitism and 'new' antisemitism are surface manifestations of a single trait, i.e. antisemitism, but that different individuals may nonetheless have a greater tendency towards one manifestation or the other. This finding provides a theoretical advance with regard to earlier studies which had merely found that the two manifestations were correlated, thus leaving open the possibility that they might represent distinct but related psychosocial traits (for the classic study, see Kaplan/Small 2006).

Secondly, it has been possible to use the data to identify the major predictors of antisemitism. The relationship between antisemitism and conspiracy thinking is well established on a historical level (see e.g. Byford 2011), but we were able to test for the strength of the relationship between antisemitism and multiple forms of conspiracy ideation, in order to identify which was the strongest.

Moreover, while a relationship between antisemitism and authoritarianism has already been found by earlier researchers (Frindte et al. 2005), we used a range of established measures to look at multiple forms of authoritarianism, establishing which were the strongest predictors.

Thirdly, it has been possible to compare the predictors of antisemitism to those of anti-Muslim prejudice. The analysis involved is complex and politically sensitive, and as such must be reported with extreme caution at this stage. However, it appears to indicate that the predictors of both forms of prejudice are very different from one another.

Bibliography

- ADL (Anti-Defamation League), 2018. Quantifying Hate: A Year of Anti-Semitism on Twitter. https://www.adl.org/resources/reports/quantifying-hate-a-year-of-anti-semitism-on-twitter (last accessed on 10 January 2021).
- ADL (Anti-Defamation League), 2019. Gab and 8chan: Home to Terrorist Plots Hiding in Plain Sight. https://www.adl.org/resources/reports/gab-and-8chan-home-to-terrorist-plots-hiding-in-plain-sight (last accessed on 10 January 2021).
- ADL (Anti-Defamation League), 2020a. Coronavirus Crisis Elevates Antisemitic, Racist Tropes. https://www.adl.org/blog/coronavirus-crisis-elevates-antisemitic-racist-tropes (last accessed on 18 January 2021).
- ADL (Anti-Defamation League), 2020b. Soros Conspiracy Theories and The Protests: A Gateway to Antisemitism. https://www.adl.org/blog/soros-conspiracy-theories-and-the-protests-a-gateway-to-antisemitism (last accessed on 18 January 2021).
- ADL (Anti-Defamation League), 2020c. Racist, Extremist, Antisemitic Conspiracies Surround Coronavirus Vaccine Rollout. https://www.adl.org/blog/racist-extremist-antisemitic-conspiracies-surround-coronavirus-vaccine-rollout (last accessed on January 25th, 2021).
- Allington, Daniel, 2018. 'Hitler had a valid argument against some Jews': Repertoires for the denial of antisemitism in Facebook discussion of a survey of attitudes to Jews and Israel. In: Discourse, Context & Media, Vol. 24, 129–136. https://doi.org/10.1016/j.dcm.2018.03.004 (last accessed on 10 January 2021).
- Allington, Daniel/Hirsh, David, 2019. The AzAs (Antizionist Antisemitism) scale: measuring antisemitism as expressed in relation to Israel and its supporters. In: Journal of Contemporary Antisemitism, Vol. 2 (2), 43–51.
- Allington, Daniel/Joshi, Tanvi, 2020. 'What others dare not say': an antisemitic conspiracy fantasy and its YouTube audience. In: Journal of Contemporary Antisemitism, Vol. 3 (1), 35–54.
- Barna, Ildikó/Knap, Árpád, 2019. Antisemitism in Contemporary Hungary: Exploring Topics of Antisemitism in the Far-Right Media Using Natural Language Processing. In: Theo-Web, 18(1), 75–92. https://doi.org/10.23770/tw0087 (last accessed: 13 January 2021).
- Becker, Matthias J., 2018 (2021). Analogien der "Vergangenheitsbewältigung": Antiisraelische Projektionen in Leserkommentaren der Zeit und des Guardian. Baden-Baden: Nomos. English title: Antisemitism in Reader Comments: Analogies for Reckoning with the Past. London: Palgrave Macmillan.
- Becker, Matthias J., 2019. Understanding Online Antisemitism: Towards a New Qualitative Approach (2019). In: Fathom: For a deeper understanding of Israel and the region. http://fathomjournal.org/understanding-online-antisemitism-towards-a-new-qualitative-approach (last accessed on 10 January 2021).
- Becker, Matthias J., 2020. Antisemitism on the Internet: An Underestimated Challenge Requiring Research-Based Action (2020). In: Justice (64) (The International Association of Jewish Lawyers), 32–40. https://www.ijl.org/justicem/no64/index.html (last accessed on 10 January 2021).
- Bortz, Jürgen/Döring, Nicola, 2006. Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer.

- Bundesamt für Verfassungsschutz, 2020. Lagebild Antisemitismus. https://www.verfassungsschutz.de/download/broschuere-2020-07-lagebild-antisemitismus.pdf (last accessed on January 25th, 2021).
- Byford, Jovan, 2011. Conspiracy theories: a critical introduction. Basingstoke: Palgrave Macmillan.
- CST (Community Security Trust), 2019. Engine of Hate: the online networks behind the Labour Party's antisemitism crisis. https://cst.org.uk/news/blog/2019/08/04/engine-of-hate-the-online-networks-behind-the-labour-partys-antisemitism-crisis (last accessed on 10 January 2021).
- CST (Community Security Trust), 2020. Coronavirus and the plague of antisemitism. https://cst.org.uk/news/blog/2020/04/08/coronavirus-and-the-plague-of-antisemitism (last accessed on 10 January 2021).
- Daniels, Jessie, 2020. The algorithmic rise of the 'alt-right'. Contexts, Vol. 17 (1), 60–65.
- Ducol, Benjami/Bouchard, Marti/Davies, Garth/Ouellet, Marie/Neudecker, Christine, 2016. Assessment of the state of knowledge: Connections between research on the social psychology of the Internet and violent extremism (TSAS Working Paper Series).
- Frindte, Wolfgang/Wettig, Susan/Wammetsberger, Dorit, 2005. Old and new anti-semitic attitudes in context of authoritarianism and social dominance orientation: two studies in Germany. In: Peace and Conflict: Journal of Peace Psychology 11 (3), 239–266.
- Hoffmann, Dagmar/Krotz, Friedrich/Reißmann, Wolfgang (eds.) 2017. Mediatisierung und Mediensozialisation: Prozesse - Räume – Praktiken. Wiesbaden: Springer VS.
- Jikeli, Günther/Cavar, Damir/Miehling, Daniel, 2019. Annotating antisemitic online content. Towards an applicable definition of antisemitism. https://arxiv.org/abs/1910.01214 (last accessed on 10 January 2021).
- Julius, Anthony, 2010. Trials of the Diaspora. A History of Anti-Semitism in England. Oxford: Oxford University Press.
- Kaplan, Edward H./Small, Charles A., 2006. Anti-Israel Sentiment Predicts Anti-Semitism in in Europe. In: Journal of Conflict Resolution, Vol. 50 (4), 548–561.
- Kneuer, Marianne/Salzborn, Samuel, 2016. Digitale Medien und ihre Wirkung auf politische Prozesse. In: Kneuer, Marianne/Salzborn, Samuel (eds.), 2016: Web 2.0 – Demokratie 3.0. Digitale Medien und ihre Wirkung auf demokratische Prozesse, Zeitschrift für Vergleichende Politikwissenschaft, Sonderheft 7, 1–15.
- Lapowsky, Issie, 2020. How COVID-19 helped — and hurt — Facebook's fight against bad content. Protocol, 11 August. https://www.protocol.com/covid-facebook-content-moderation (last accessed on 31 January 2021).
- Mayring, Philipp, 2015. Qualitative Inhaltsanalyse. Grundlagen und Techniken. Weinheim und Basel: Beltz.
- Meibauer, Jörg, 2008. Pragmatik. Eine Einführung. Tübingen: Stauffenburg.
- Müller, Karsten/Schwarz, Carlo, 2020. Fanning the Flames of Hate: Social Media and Hate Crime. http://dx.doi.org/10.2139/ssrn.3082972 (last accessed on 19 January 2021).
- Nagle, Angela, 2017. Kill All Normies: Online Culture Wars from 4chan and Tumblr to Trump and the Alt-Right. Winchester: Zero Books.
- Ozalp, Sefa/Williams, Matthew L./Burnap, Pete/Liu, Han/Mostafa, Mohamed, 2020. Antisemitism on Twitter: Collective Efficacy and the Role of Community Organisations in Challenging Online Hate Speech. https://orca.cf.ac.uk/132742/1/2056305120916850.pdf (last accessed on 10 January 2021).

- Rich, Dave, 2016. The Left's Jewish Problem: Jeremy Corbyn, Israel and Anti-Semitism. London: Biteback.
- Rieger, Diana, 2019. Diskussionsräume und Radikalisierungsprozesse in sozialen Medien. https://www.bpb.de/politik/extremismus/rechtspopulismus/290851/diskussionsraeume-in-sozialen-medien (last accessed on 19 January 2021).
- Roose, Jochen 2020. Verschwörung in der Krise. Repräsentative Umfragen zum Glauben an Verschwörungstheorien vor und in der Corona-Krise. Berlin: Konrad-Adenauer-Stiftung. https://www.kas.de/documents/252038/7995358/Verschw%C3%B6rung+in+der+Krise+%28PDF%29.pdf/7703c74e-acb9-3054-03c3-aa4d1a4f4f6a?version=1.1&t=1608644973365 (last accessed on January 25th, 2021).
- Schoeps, Julius H./Schlör, Joachim (eds.), 1996. Antisemitismus. Vorurteile und Mythen. Munich: Piper.
- Schwarz-Friesel, Monika/Reinharz, Jehuda, 2017. Inside the Antisemitic Mind: The Language of Jew-Hatred in Contemporary Germany. Brandeis: Brandeis University Press.
- Schwarz-Friesel, Monika, 2019. Antisemitism 2.0: The Spreading of Jew-hatred on the World Wide Web. In: Lange, Armin/Mayerhofer, Kerstin/Porat, Dina/Schiffman, Lawrence H. (eds.), 2019. An End to Antisemitism! Volume 1: Comprehending and Confronting Antisemitism. New York: De Gruyter, 311–338. https://doi.org/10.1515/9783110618594-026 (last accessed on 19 January 2021).
- Shahrezaye, Morteza/Meckel, Miriam/Steinacker, Léa/Suter, Viktor, 2021. COVID-19's (mis)information ecosystem on Twitter: How partisanship boosts the spread of conspiracy narratives on German speaking Twitter. Future of Information and Communication Conference (FICC). Vancouver, Canada.
- Stegemann, Patrick/Musyal, Sören, 2020. Die rechte Mobilmachung. Wie radikale Netzaktivisten die Demokratie angreifen. Berlin: Econ.
- Thomas, Zoe, 2020. Facebook content moderators paid to work from home. BBC News, 18 March. https://www.bbc.co.uk/news/technology-51954968 (last accessed on 30 January 2021).
- Williams, Matthew L./Burnap, Pete/Javed, Amir/Liu, Han/Ozalp, Sefa, 2020. Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime, The British Journal of Criminology, Volume 60 (1), 93–117. https://doi.org/10.1093/bjc/azz049 (last accessed on 19 January 2021).
- Woolley, Samuel/Joseff, Katie, 2019. Jewish Americans. Computational propaganda in the United States: Trends in anti-Semitic harassment and political disinformation on social media. Institute for the Future. https://www.iftf.org/fileadmin/user_upload/downloads/ourwork/IFTF_JewishAmerican_comp.prop_W_05.07.19.pdf (last accessed on 10 January 2021).
- YouTube Team, 2020. Responsible policy enforcement during Covid-19. YouTube official blog, 25 August 2020. https://blog.youtube/inside-youtube/responsible-policy-enforcement-during-covid-19 (last accessed on 30 January 2021).
- Zannettou, Savvas/Finkelstein, Joel/Bradlyn, Barry/Blackburn, Jeremy, 2020. A Quantitative Approach to Understanding Online Antisemitism. In: Proceedings of the International AAAI Conference on Web and Social Media, 14(1), 786–797. https://ojs.aaai.org/index.php/ICWSM/article/view/7343 (last accessed on 10 January 2021).