האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

בית ספר פדרמן למדיניות ציבורית וממשל
FEDERMANN SCHOOL OF PUBLIC POLICY & GOVERNANCE
مدرسة فديرمان للسياسة العامة والحكم

The Autonomy of Regulatory Intermediaries:

Meta and The Oversight Board as a Case Study

Omer Shiloni

Advisor: Prof. David Levi-Faur

A Dissertation Submitted to the Federmann School of Public Policy & Governance,

The Hebrew University of Jerusalem

December 2022

Jerusalem

Abstract

This paper examines the credibility of the Meta created Oversight Board as a solution to content moderation issues on online social media platforms. It focuses on its autonomy as a self-regulation model of corporate created regulatory intermediary and the dual relationship it has with Meta. The paper aims to answer two questions: (1) What are the condition under which a corporate created regulatory intermediary can operate with a large enough degree of autonomy to become a reliable critic with the ability to affect actions and policies of said corporate? and (2) What outside- and/or inside-factors influence Meta's willingness to comply with the Board non-binding recommendations? To answer the first question, I analyze the Oversight Board institutional design and its de-facto autonomy, using various empiric measures of autonomy to assess the unique case of the Oversight Board. This analysis indicates that the Board was created with a high degree of official autonomy and can operate with a high degree of de-facto autonomy, and thus can operate as a reliable critic. To answer the second question, I use quantitative analysis to assess the influence of outside factors, namely press coverage and stock price, and qualitative analysis, based on interviews with Board personnel, media interviews of Board members and public documents, to assess inside factors, namely the Oversight Board's relationship with Meta and the strategies it employs to garner respect and bolster its autonomy. The quantitative analysis found no significant correlation between Meta's press coverage or stock price and its willingness to comply with the Board's recommendation. The qualitative analysis identified that the Oversight Board developed a self-image of an independent autonomous entity not beholden to Meta which engages in developing high-end policy solutions to content moderation. This self-image is supported by three practical strategies: (1) acts intend to establish the Board as a professional and impartial entity, (2) developing a communication method with Meta which gives the Board access to relevant information while at the same time prevents interference from Meta and (3) engaging in direct acts of self-assertion. These strategies partially correlate to known strategies used by interdependent organizations to boost respect and autonomy. This paper concludes that the Oversight Board can operate with a high degree of autonomy, and that the strategies it employs influence Meta's compliance with its recommendations, and therefore presents a viable solution to content moderation issues, as well as a model for other industries striving for efficient self-regulation.

Acknowledgments

It is my pleasant duty to thank the following people, without whom this paper would not have been possible: To my advisor, Prof. David Levi-Faur, for his guidance, support, patience and encouragement; to Dr. Dmitry Epstein, for friendship and for always finding time to listen and help; to Prof. Sharon Gilad, for very sage advice, several of them actually; to Dr. Liat Raz-Yurovich, for teaching me how to love statistics and not be afraid of code; to Prof. Raanan Sulitzeanu-Kenan, for a very illuminating 45 minutes about Anova, Regression and Poisson; to Dr. Nir Rotem, for guiding the initial research as part of his course on qualitative methods; and to Dr. Tehilla Shwartz Altshuler, for getting me into this mess.

This paper is dedicated to my partner, Anat, and my sons, Jonathan and Geffen. You are the reason for everything.

Table of Contents

List of Tables and Figures

1. Introduction

In November 2018, bowing to growing criticism, Mark Zuckerberg, founder and CEO of Meta Platforms (then still known as Facebook) approved a bold idea: the creation of a new external body which will serve as an arbiter of the company's content moderation decisions. That body, *the Oversight Board*, sometimes nicknamed "Facebook's Supreme Court" (Klonick, 2021), founded and funded by Meta, was intended to operate as an autonomous entity, not owned or directly controlled by Meta. The Board's mandate authorizes it to make binding rulings in specific cases where a user of Facebook or Instagram (platforms owned by Meta) or Meta itself have asked it to review a content moderation decision. The Board can also present broader policy advisories, or recommendations, to which Meta is obligated to reply but not implement. The decision to create the Board was received with skepticism, with some commentators declaring it "destined to fail", "weak", "toothless" or simply "not enough", sometimes mere hours after it was announced (Constine, 2020; Ghosh, 2019; Hensel, 2018).

This paper offers an in-depth examination of the Oversight Board, assessing its reliability as a solution to content moderation problems faced by big social media platforms, through the theoretical framework of regulatory intermediaries. Its goal is to explain the theoretical and practical framework in which the Oversight Board operates; evaluate the Board's operation and decision-making based on new and existing models; assess its influence on Meta during its first months of operation and the factors behind it; analyze how Oversight Board members and staff perceive their role in shaping Meta's policy decision; and explore the strategies employed by the Board to boost its autonomy.

To date, no such in-depth attempt to methodically evaluate the Oversight Board's has been made. Current academic research regarding the Board is scant, and mostly examines the entity through a legal perspective. The earliest scholarly paper, by Evely Douek of Stanford Law, was published less than a year after the Board was created, and almost a year before starting operations. That paper evaluates the Board's potential based on then available information, and deals mostly with mapping Meta's vision for the Board, its expected limitations and the ways Meta can maximize the Board's potential (Douek, 2019). Another noted article, from the Yale Law Review, minutely details the Board's creation process, briefly reflects on some aspects of the Board's independence based on its charter and aggregates various views on its possible influence (Klonick, 2020). A later article focuses on the claim that the Board's charter gives it power to review Meta's algorithms, an endeavor the Board is yet to attempt (Pickup, 2021).

Articles examining on the Board's actual operation and decision-making first appeared in the middle of 2021. In the first of which, the author provides a selective and limited analysis of the Board's first few decision, focusing on perceived problems such as use of public law narrative and a bias towards freedom of speech arguments (Schultz, 2021). An article written in the wake of the Board's decision in the case of Donald Trump's suspension discusses its place in Meta's ecosystem of relationships with states, publics and staff (Arun, 2022). Two latter articles from Brenda Dvoskin mostly focus on the of framework International Human Rights Law, with one analyzing how the Board uses this framework to build objectivity and legitimacy for its decisions, decrying it as an ineffective tool creating a so-called "expert governance" and suggesting reliance on promotion of civil society involvement instead (Dvoskin, 2022a, 2022b). A more recent study uses the Board's decisions the map its perceived strengths (transparency of content moderation, influential policy recommendations and assertiveness) and weakness (limited jurisdiction, limited impact, Meta's control over precedent and lack of diversity), illustrates them through an analysis of a Board decision, and ends with presenting four recommendations to improve the Board (D. Wong & Floridi, 2022).

All previous research was done by law-scholars, focused on legal or judicial aspects of the Oversight Board's operation, and did not have much operational data to rely on. To date, no systematic and in-depth approach has been applied to examine the Board's autonomy, operations, decision-making and influence on Meta, nor has such an examination been conducted from a public policy perspective. This paper aims to examine the Board's autonomy, operation and influence on Meta, using a multi-method approach utilizing modelling, quantitative and qualitative analysis based on known and new metrics and incorporating publicly available information as well as in-depth interviews with Oversight Board personnel.

The paper proposes two answer two questions: (1) What are the condition under which a corporate created regulatory intermediary can operate with a large enough degree of autonomy to become a reliable critic able to affect the actions and policies of said corporate? And (2) what external- and/or internal-factors influence Meta's willingness to comply with the Board's non-binding recommendations?

This paper proceeds in three parts. Part II lays the practical and theoretical foundation for the research, with an overview of the Oversight Board, discussion of regulation of large online platforms and content moderations and a literary overview of current research into regulatory intermediaries and self-regulation.

Part III explores the formal autonomy of the Oversight Board, utilizing the research framework around formal independence, and adapting well-known indices developed to assess the political independence of regulatory agencies to use in the case of the Board; as well as the de facto autonomy of the Board by assessing the level of burden Meta will incur in implementing the Board's recommendations.

Part IV evaluates the impact the Oversight Board had on Meta, and the factors behind it. Seeing that Meta accepts most the Board's recommendation, I use quantitative and qualitative analysis (the latter based in part on in-depth interviews with several Board personnel) to determine whether Meta is influenced by external factors in deciding to accept the Board's recommendations, or whether it was the actions of the Board itself, employing tactics used to garner respect and autonomy, that can explain Meta's acceptance.

2. Practical and Theoretical Framework

2.1 The Oversight Board – a Brief Overview

The Oversight Board's two governing documents are the Board's charter and bylaws. The charter "specifies the board's authority, scope and procedures", and establishes an "irrevocable trust with trustees" to manage the Board's operation (*Oversight Board Charter*, 2019). Meta endowed the trust with $130 million at its creation in 2019, and endowed a further $150 in 2022 (Oversight Board, 2022c). The charter establishes the Oversight Board as a three pronged organization: (1) The trust and trustees, which control and manage the Board's budget and administration, and are appointed directly by Meta; (2) The Board members, numbering up to forty and headed by four co-chairs, who select, discuss and issue decisions on case; and (3) the Oversight Board administration, headed by a director (a role equivalent of a CEO) appointed by the trustees and staffed by a full-time staff hired by the director or underlings. Though administrative staff provides support for the Board members in case selection, research, discussion and decision-making, they do not report directly to Board members but rather to the Trust. As one Board staff stated in an interview: "The board members definitely are not our bosses (…) it is the trustees who could fire or hire Thomas [the Board's director at the time of interview; O.S.]. So theoretically, I think the trustees are probably our bosses".

The Charter also establishes five powers of the Oversight Board: (1) request information from Meta required in its deliberations, (2) interpret Meta's relevant content policies, (3) instruct Meta to remove or preserve content and to preserve or overturn a designation which led to an enforcement outcome, (4) issue written decisions and (5) provide non-binding policy advisories (*Oversight Board Charter*, 2019).

The Oversight Board's bylaws "specify the operational procedures of the board" (*Oversight Board Charter*, 2019), and unlike the charter can be amended at the Board's discretion. The bylaws stipulate the Board's case selection and membership selection process (both through a committee of Board members which rotates regularly), and the case deliberation process: first by a panel of five Board members, which discuss, order research, consults with experts, Meta representatives and other stakeholders, and writes a draft decision. This draft is then reviewed by the whole Board, which can order the panel the re-review the decision, and finally approve and present the final decision (Oversight Board, 2022d).

The Oversight Board started operation in October 2020 and its first five rulings were published in January 2021 (Oversight Board, 2020, 2021a). In four of the cases, the Board has overturned Meta's previous decisions. More noteworthy than the Board's rulings were the policies and theoretical frameworks developed through its policy advisories, meant to guide and shape Meta's policies and operations. These include a call for more transparency when engaging with users about content moderation decisions, and an innovative approach to content moderation. This approach states that decisions should be based not on a rigid set of rules, but on a dynamic interpretation of how said content is viewed by its author and its intended audience, with the specific time and place it was created and the relevant culture in which it's consumed as a point of reference (*Case Decision 2020-002-FB-UA*, 2021; *Case Decision 2020-003-FB-UA*, 2021; *Case Decision 2020-004-IG-UA*, 2021; *Case Decision 2020-005-FB-UA*, 2021; *Case Decision 2020-006-FB-FBR*, 2021).

2.2 Regulatory Intermediaries as a Framework to the Regulation of Online Platforms

The so-called five big tech companies – Apple, Google, Meta, Amazon and Microsoft – present a diverse set of challenges to regulators. In addition to antitrust issues common to regulating big multi-national corporations, big tech presents idiosyncratic challenges such as worldwide platforms operating in many jurisdictions, wide dispersion of operation and audience, and the sheer size and complexity of the legal entities, each encompassing hundreds of thousands of employees serving billions of users and operating through a convoluted network of subsidiaries spread across dozens of jurisdictions. Repeated attempts to regulate big tech companies, and specifically online platforms, on issues such as content moderation, privacy and use of data, have proved challenging, partially since most policies applied to platforms were crafted before their emergence, address a broader category of online services, and weren't created with the present issues in mind (Budzinski & Mendelsohn, 2021; Smyth, 2020).

Content moderation presents an especially salient issue when discussing online platforms, and recent years have demonstrated that failure to regulate online content can result in significant harm in the offline world. Major events, such as the genocide of the Rohingya people in Myanmar, the civil war in Ethiopia, the results of the 2016 US presidential elections and the January 6th riots at the US Capitol have all been linked to content moderation failures of online platforms, Meta's Facebook prominent among them (Amnesty International, 2022; Dutt et al., 2019; Kurtzleben, 2018; Mackintosh, 2021; Milmo, 2021; Ng et al., 2022; Robins-Early, 2021). The main rule regulating content moderation on user generated online platforms is section 230 of the United State Communications Decency Act of 1996, which states that online platforms will not be considered as publishers or speakers vis-à-vis content created by third parties, such as users, thereby protecting them from suits or liability related to that content (Communications Decency Act of 1996, 1996). The law also allows platforms to remove or moderate third party content if they deem it obscene, inciting to violence, harassing or otherwise objectionable, even if it is protected by the First Amendment to the United State Constitution. This legal framework contributed to social media platforms being perceived not as content producers but as intermediaries (Gillespie, 2017), and enabled the development of content moderation regime based mainly on voluntary self-regulation, motivated mostly by desire to maintain an environment hospitable to the advertisers (Gillespie, 2018). Meanwhile, voluntary self-regulation of content by the platforms have raised concerns of sidelining state regulation, usurping existing institutions and enabling private censorship, while avoiding governmental oversight and delegating content moderation decision to under-regulated machine learning algorithms (Grabosky, 2013; Langvardt, 2017; Medzini, 2021b; Yeung, 2018). Gaps between rule-makers and the general public in the understanding of central regulatory issues such as privacy further complicate the issue (Epstein & Medzini, 2021).

One possible under-represented solution to regulate big tech companies might be through regulatory intermediaries. Regulatory intermediaries are independent bodies or experts that provide external assistance to regulators in achieving their goal of regulating rule-takers, thereby creating a process in which regulation is implemented using mediating activities (Abbott et al., 2017a; Brès et al., 2019). Where the typical model of regulation deals with a two actors relationship – rule-maker (R) and rule-taker or target (T), and can be represented as R→T – the use of regulatory intermediaries (I) requires a three- or molti-actors relationship: R→I→T (Abbott et al., 2017a). They can come from the private sector, i.e., accounting firms or rating agencies, the third sector, and even government agencies and countries in some cases (Abbott et al., 2017a; Levi-Faur & Starobin, 2014). Regulatory roles of intermediaries include reporting, auditing, ranking and certification, and they might also serve in an expert or counselor capacity (Kourula et al., 2019). Their function is not restricted to activities of state or regulatory agencies, and can include other forms of regulation: public, private and hybrid; national, international and transnational; formal and informal (Abbott et al., 2017b). Notable examples of use of regulatory intermediaries include private entities certifying Kosher food products, credit rating agencies such as Moody's and S&P (Abbott et al., 2017b), entities labeling the energy consumption of electrical devices, independent ranking of higher education institutes, the European Union (EU) use of national agencies'' transgovernmental networks to consistently implement rules and regulations (Abbott et al., 2017a; Levi-Faur, 2011) the use of The United States Food and Drug Administration (FDA)of private auditors to monitor food imports (Lytton, 2017), the work of International Criminal Court with NGOs to enlist the cooperation of various governments (De Silva, 2017), and even lawyers, accountants, investment bankers and inspectors (Levi-Faur & Starobin, 2014). One role a regulatory intermediary might take, and which is especially salient to the case of the Oversight Board, is promoting the implementation of rules through interpreting and elaborating them, in essence translating them for practical use (Abbott et al., 2017b). In arbitrating cases, the Board must interpret Meta's own content moderation policies, adapting them to relevant situations. Through its policy advisories, the Board also take part in shaping and developing those policies.

In typifying regulatory intermediaries, a distinction may be made across two dimensions: official/unofficial and formalized/non-formalized (Brès et al., 2019). In the first dimension, official is defined as decreed or legislated by a legitimate authority and unofficial as intermediation outside the mandate of such authority. In the second dimension, formalization is the process of “turning tacit processes into explicit and ‘material’ ones” (Brès et al., 2019, p. 130). Combining these two dimensions produces four types of regulatory intermediaries: (1) formal (official and formalized), where an official authority endorses an intermediary and delegates tasks for enforcement or monitoring; (2) interpretive (official and unformalized), where an intermediary is endorsed by an official entity, while its processes are largely undefined; (3) alternative (unofficial and formalized), a well-organized intermediary operating outside and even against official regulation; and (4) emergent (unofficial and unformalized), an unexpected intermediary with the capacity to control and affect relations between rule-makers and rule-takers (Brès et al., 2019; Levi-Faur & Starobin, 2014). In examining the Oversight Board, however, it is apparent it does not fit easily into one of the four types: Meta is not an official government entity, thereby landing an unofficial aspect to the Oversight Board’s creation. On the other hand, Meta is the authority empowered (by laws and custom) to enforce content moderation policy on its platforms, which lands the Board a more official aspect. And while the Board’s operation is highly formalized, through its charter and bylaws, the Board does engage in expanding its roles in practice, and only later, if at all, codifying them by amending its bylaws. For example: the Board’s unilateral decision to monitor and report on Meta’s implementation of its recommendations (Oversight Board, 2022b). This indicates that a more nuanced understanding of regulatory types is warranted, one that can accommodate a flexible and complex system where an entity might be both rule-taker and rule-maker, and where the role of the intermediary itself is in constant flux.

Another aspect of regulatory intermediaries is the complication they introduce to the basic principal-agent model, being a form of regulation-by-delegation (Medzini & Levi-Faur, 2022).

Reasons for delegation of responsibilities to intermediaries are varied and might encompass political interests, capacities the delegator lack but are present at the intermediary, or a will to influence other actors behaviors (Medzini, 2021a, 2021b; Medzini & Levi-Faur, 2022). Under the principal-agent model, the principal delegates to reduce costs or increase commitments to policy (Majone, 2001; Medzini & Levi-Faur, 2022). In a basic principal-agent model, the focus is on direct relationship between the actors (Abbott et al., 2017a). The introduction of intermediary shifts the focus to a polycentric relationship, where an intermediary itself might relay on other intermediaries (Medzini & Levi-Faur, 2022).

Intermediaries expand the scope of regulatory analysis, incorporating actors beyond the traditional rule-maker and rule-takers, enabling a more complex and nuanced understanding of the ways in which regulation is implemented and even created, and how rule-takers interact, directly or indirectly, with rule-makers by creating, selecting and engaging with intermediaries (Abbott et al., 2017a). Examining the way regulatory intermediaries operate and the effect they have on both rule-makers and rule-takers is therefore crucial for a better understanding of regulatory governance, and helps identify the way regulatory intermediaries shape relationship between rule-makers and rule-takers, and the various direct and indirect mechanisms behind it (Abbott et al., 2017a).

Regulatory intermediaries have an enhanced role in the framework of transnational regulatory governance, where corporations operate across different countries, and are not under the authority of one government. This framework is considered more diverse and messier than traditional regulatory approaches at the national level, though similar in principal (Levi-Faur & Starobin, 2014). From the viewpoint of regulators, employing intermediaries can enhance the regulation process by incorporating new capacities or expanding existing ones. These include a greater operational capacity in areas where a regulator lacks capacity or the resources to employ it; expertise not readily available to a regulator; and legitimacy, flowing from the reputation of the intermediary (Abbott et al., 2017a).

The importance of regulatory intermediaries comes to light when they fail to fulfill their role. The 2007–2008 global financial crisis was caused, in part, when credit rating agencies, favoring profits over integrity, assigned diminished risk values to risky financial devises, thereby contributing to the creation of a real-estate bubble in the US market (Abbott et al., 2017a). A deadly 2012 fire in a Karachi factory occurred shortly after the factory was certified by a private auditor, raising questions as to the reliability of private regulatory intermediaries (Levi-Faur & Starobin, 2014). Cases such as these serve to undermine the legitimacy, autonomy and regulatory capture of regulatory intermediaries, necessitating a closer look at their institutional design and de facto operation (Abbott et al., 2017a).

Regulatory intermediaries may also be considered as a form of self-regulation. Self-regulation can be described has the set of rules and norms designed and managed by private entities, at the corporation- or industry-level and intended to constrain the conduct of said entities (Porter & Ronit, 2006). Corporations or industries might turn to self-regulation in an effort to impede or supplement government rules and regulations governing their activities (Cusumano et al., 2021). In current literature, there is no agreed definition of self-regulation, and the term has been described as extremely malleable, able to encompass a variety of instruments (Cusumano et al., 2021; Sinclair, 1997). One common definition for self-regulation is any rule created and enforced by a non-governmental actor, with a more narrow one defining it as rule created and enforced by the regulated entity (Coglianese & Mendelson, 2010; Cusumano et al., 2021).

Gunningham & Rees (1997), disputing the uniformity of unnuanced self-regulation definitions, offer several distinctions in typifying self-regulation. The first, distinguishes between individual self-regulation and self-regulation by group. Individual self-regulation or self-regulation at the corporation level, might include self-monitoring of regulatory violations, proactive initiatives to improve public image, voluntary agreements, environmental covenants and negotiated compliance (Cusumano et al., 2021; Sinclair, 1997). At the group or industry level, self-regulation can be set and enforced by an voluntary association of companies operating in the same market through an agreed code of practice (Gunningham & Rees, 1997); such as the Motion Picture Association of America (MPAA), which maintains a movie rating system and in the past enforced the now notorious Motion Picture Production Code (commonly referred to as the Hays Code), dictating content standards for film productions (A. J. Campbell, 1998).

A second distinction is made between economic self-regulation – i.e., control of markets and other aspects of economic life – and social self-regulation – i.e., protecting people or the environment from damages caused by industrialization (Gunningham & Rees, 1997; Hawkins & Hutter, 1993). The third distinction observes the degree of government involvement divided into three forms: voluntary self-regulation, in which a corporation or an industry both create and enforce rules; mandated full self-regulation where rules are created and enforced by the private entity, but unlike voluntary self-regulation the scheme is sanctioned and monitored by the government; and mandated partial self-regulation, in which the private entity either create or enforce rules, but not both (Gunningham & Rees, 1997; Rees, 1988). Some scholars described the latter as enforced self-regulation, where the regulator or government mandates firms to self-regulate through the division to regulatory rules, self-monitoring and even self-punishing for cases on noncompliance (Ayres & Braithwaite, 1992). Another notable form of self-regulation is meta-regulation, meaning the self-evaluation and betterment of self-regulation mechanisms (Gilad, 2010), or, according to another definition, the ways regulators deliberately induce targets to develop self-regulation (Coglianese & Mendelson, 2010). These two contradictory definitions exemplify the difficulty and disagreement researchers face in defining self-regulation.

In typifying the Oversight Board according to the three distinctions presented by Gunningham & Rees, one might describe it as a scheme of individual self-regulation, concerned with social self-regulation and operating a form of voluntary self-regulation. This definition is tempered by the Board's ambition to serve other platforms, creating the possibility of it becoming a scheme of industry self-regulation. The separation of the Board from Meta, as a distinct legal entity, indicates a possibility for another fourth distinction of self-regulation, which focuses on the autonomy of the entity charged with self-regulation.

Most theory around self-regulation regimes is based on a few assumptions. These include the attempts of regulatory regimes to strengthen self-regulation schemes while limiting their shortcomings, and focus on direct-hierarchical relationships between policy makers and targets or between regulators and targets, neglecting the role of the principal-agent relationship between policy makers and regulators (Ayres & Braithwaite, 1992; Héritier & Eckert, 2008; Medzini & Levi-Faur, 2022). More relevant to this paper is the lack of attention to the interaction of self-regulation models with regulatory intermediaries, which can increase the legitimacy of self-regulation (Medzini & Levi-Faur, 2022).

Supporters of self-regulation cite benefits such as speed, flexibility, sensitivity to market conditions and lower costs, thanks in part to the target's greater knowledge of its operations, and better compliance to self-created rules (Coglianese & Mendelson, 2010; Cusumano et al., 2021; Gunningham & Rees, 1997). In practice, however, self-regulation tends to fail to achieve its stated goals, and serves industry interests rather than public ones, thereby creating no more than the appearance of regulation intended to hinder direct government intervention (Gunningham & Rees, 1997). One researcher even dammed self-regulation as ”an attempt to deceive the public into believing in the responsibility of an irresponsible industry” (Braithwaite, 1993, p. 93). If the Oversight Board proves a reliable form of self-regulation, its model may serve as a template for other corporations or industries interested in employing self-regulation while avoiding the pitfalls usually associated with it. Some conditions under which self-regulation might thrive are related to the size of the industry and the homogeneity of the actors, as well as a common understanding of shared interests and the ability to monitor actions and implement sanctions (Cusumano et al., 2021). These conditions apply to self-regulation at the industry level and are therefore mostly irrelevant to the case of the Oversight Board.

“Enhanced self-regulation”, a term coined by Medzini & Levi-Faur (Medzini, 2021a, 2021b; Medzini & Levi-Faur, 2022), represents the confluence of self-regulation and regulatory intermediaries, referencing a model where organizations rely on intermediaries to restrict behavior and gain credibility. This model is contrasted with the of “thin self-regulation”, where the focus is on the interactions between regulators and target organizations, or where intermediaries are dependent and lack autonomy and credibility, or are not the primary regulatory mechanism (Medzini & Levi-Faur, 2022). In an enhanced self-regulation model, entities willingly restrict their use of power by delegating responsibilities to independent intermediaries in order to increase the credibility of their commitment to certain policies (Medzini & Levi-Faur, 2022). A model of enhanced self-regulation may incorporate specific control mechanisms and its focus is on delegation of responsibilities to third parties, chosen mixture of tools and implications to the regulatory regime and to policy (Medzini & Levi-Faur, 2022).

Currently, with the exception of extreme cases, mostly content related to support of terrorism and child sexual abuse material (CSAM), virtually all content moderation of online platforms is handled through a thin self-regulation model at the corporation level; with different firm setting different rules and utilizing different enforcement measures, most of which are not transparent or open to public critic (Hartmann, 2022). Meta main model for self-regulating content centers on a convoluted set of policies and community standards, enforced through a combination of mass-user auditing, human moderators tasked with adjudicating content and machine learning algorithms. The approach, utilized by other content platforms such as YouTube and Twitter, has been criticized by regulators, consumer advocacy organizations, researchers and even former industry insiders as inefficient, lacking and failing to achieve stated goals such as curbing the distribution of disinformation. This, in turn, led to increased calls for more government oversight of online platform content moderation policies, with some governments, most notably the European Union, responding with relevant legislation (Fernández, 2020; Redrup & Tillett, 2019; *Self-Regulation of Social Media Platforms Failing to Curb Disinformation, Says New Report*, 2019).

Other models used by Meta, more in the vain of enhanced self-regulation, include a cooperation with government entities, such as the State Attorney of Israel, which enables them to “fast track” complaints about abusive content (Kabir, 2021), and working with independent fact checkers to review and label disinformation or misinformation (Meta, 2021). Meta’s page admins & users, employed to implement self-imposed rules and to manage content issues on its platforms, might also be viewed as regulatory intermediaries, part of an enhanced self-regulation model (Medzini, 2021b). In another scheme, Twitter employed a Trust and Safety Council, which advised it on how to enforce content policies. The council differed from the Oversight Board, as it was a composed of civil society groups working voluntarily, had advisory powers only and was controlled entirely by Twitter with no staff or funding of its own. In the wake of Elon Musk takeover of Twitter, the council was disbanded (Haggin, 2022). The lack of autonomy of the council and its dependency on Twitter would typify it as a form of thin self-regulation, rather than enhanced self-regulation.

The Oversight Board represents a new approach to Meta's enhanced self-regulation practices, one that is more official and formalized, and that is geared towards addressing the challenges of regulating content on a worldwide multi-language online media platform. The Board is an uncommon institution from a regulatory intermediary perspective, as it has dual relationship with Meta, which acts both in the role of the rule-maker, as the creator of the Board, and as well as that of the rule-maker, somewhat complicating the basic R→I→T model conceived by Abbott, Levi-Faur and Snidal (2017a). This model assumes a "unidirectional flow of intention and action" (Abbott et al., 2017a, p. 17), under which the rule-maker formulates regulations, chooses intermediaries to implement the regulation at the rule-taker. In the case of the Oversight Board, however, Meta plays a dual role, suggesting a more cyclical model is in place. This model is than further complicated as the Board is not only charged with enforcing and interpreting regulations (i.e., Meta's content moderation policies), but also with criticizing existing rules and suggesting new ones, placing the Board itself in the dual role of both intermediary and rule-maker. In addition, the influence is not unidirectional, as any action the Board takes toward "Meta the rule-taker" will necessarily influence "Meta the rule-maker", while the Board's formal autonomy and institutional design limit the action Meta might take vis-à-vis the Board after its initial setup. An analysis of the Oversight Board's operation and relationship with Meta will therefore contribute to a more nuances understanding of the roles regulatory intermediaries take and their cross-influence on rule-makers and rule-takers.

Meta's use of the Oversight Board enables it to employ some of the extended capacities intermediaries provide to regulators, notably legitimacy and expertise, through the scholars, researchers, human rights activists, journalist and other prominent figures serving as Board members. Legitimacy, however, can only be achieved with the Oversight Board having a large degree of autonomy. In general, autonomy is the ability of an actor to operate freely and without hindrance from other entities. Autonomy must not only be granted by parties that might interfere with an actor actions, but must also be claimed by the actor through autonomous actions (Wiedner & Mantere, 2019). The autonomy of a regulatory intermediary can be assessed by its ability to translate "its own preferences into authoritative actions" (Nordlinger, 1987, p. 355). A wholly autonomous body will act, or will not act, solely according to its own interests and goals (Nordlinger, 1987). For a regulatory intermediary to operate effectively, it must enjoy a autonomy vis-à-vis the rule-maker and the rule-taker (Abbott et al., 2017a). Notably, autonomy vis-à-vis the rule-taker is important to achieve legitimacy.

The dual relationship of the Oversight Board with Meta requires the assessment of two form of autonomy: political- or formal-autonomy from Meta in its role as the rule-maker, and de facto autonomy from Meta in its role as the rule-taker, the latter of which can also be viewed as a measure of regulatory capture. Regulatory capture may be defined as the process which enables rule-takers to disrupt intervention by a regulatory agency, or the process where rule-takers manipulate the regulatory agency appointed to control them (Dal Bó, 2006). Stigler (1971) suggest that regulation is designed to primarily benefit the rule-takers. In this view, regulatory capture is the result of stakeholder's demand to initiate regulation (Dal Bó, 2006; Stigler, 1971). When discussing regulatory intermediaries, regulatory capture might manifest in its classical form – where the target dominates the regulator, or in a more nuanced form in which the intermediary is captured by the regulator, thereby creating an illusion of action to divert from a regulators failures (Abbott et al., 2017a, 2017b) – making capture another aspect of autonomy. This approach is especially salient in understanding the relationship between Meta and the Oversight Board, as the first is the creator of the latter.

Meta owns and operates the world's most popular inter-personal media platforms. Its platforms – Facebook, Instagram, WhatsApp and Messenger – have 3.71 billion monthly active users (MAUs) and Facebook alone has 2.96 MAUs, according to its October 2022 quarterly report (Meta, 2022). It has an unprecedented influence on human discourse and world events and can impact a large swath of the world's population and most of the western world's perception of reality. Yet its operation is maligned with problems ranging from privacy violations, rampant hate speech, misinformation, violence and extremism, and repeated failures to ensure users' safety. These problems are not unique to Meta and are typical of other social media platforms. Understanding and analyzing the operation of the Oversight Board as a possible remedy to address these issues, will provide a better understanding of which regulatory solutions work and can be applied successfully to other organizations and other industries, and which are not reliable.

Examining and understanding the operation of the Oversight Board as a measure to mitigate content moderation issues will provide important insights to this new approach; its reliability as an effective solution to other social media platforms; and its possible influence on measure currently developed or considered by decision-makers. It will also shed a light on the processes used by multi-national organizations to develop regulatory solutions and will help expand the theory on regulatory intermediaries through the examination of a new type of regulatory intermediary.

This paper will also help expand the theoretical understanding of regulatory intermediaries, by dissecting a new under-studied model of a regulatory intermediary – one created and funded by a private entity, for to sole purpose of regulating the very same entity. This examination will improve the knowledge of the interaction, cross-influence and interdependency of regulatory intermediaries. It will provide insights into the efficacy of alternative models of self-regulation, with implications not only to content moderation issues, or other issues facing big platforms in general, but also to other industries facing challenges with the applications of self-regulation. For example: the chemical manufacturing industry operates a voluntary self-regulation scheme at the industry level, maintained by the trade association The American Chemistry Council (ACC). The scheme, Responsible Care Program, aims to reduce pollution from the manufacturing of chemicals through submission of annual reports on progress towards code implementation and sharing of pollution abatement information (Gamper-Rabindran & Finger, 2013). However, researchers criticized the scheme as enabling opportunism and ineffective, with findings noting that factories enrolled in the scheme produce more pollution then those that are not (Gamper-Rabindran & Finger, 2013; King & Lenox, 2000). The insights from the current research can be useful for industries such as this and provide a pathway for a better model of self-regulation.

I theories that the Board's members see its role not as that of an arbiter in specific cases, but as an autonomous body engaged of high-level policy solutions, whose role is to provide a theoretical, practical and moral roadmap for Meta's content moderation practices. Tension and conflict between the Board and Meta serves as catalysts for the Board to assert itself as an autonomous influential voice in guiding Meta's policy decisions. This, in turn, might offer a reliable model of enhanced self-regulation to other social media platforms engaging in content moderation decisions.

2.3 Methodology

In this paper, I use various research methods to assess the reliability of the Oversight Board as a regulatory intermediary. Each method is described thoroughly in the relevant section. Here follows a brief overview of the methods used. The autonomy of the Board is evaluated through the research framework of independence of regulatory agencies. To assess the Board's formal autonomy, I use modified versions of indices created by Gilardi (2002, 2005) and expanded by Hanretty & Koop (2012; Koop & Hanretty, 2018). The indices were created to evaluate the independence of regulatory agencies from governments and/or legislators but can be modified to assess the formal autonomy of the Board from Meta, as its creator and funder, taking the place of government. The current research on de facto independence, mostly done by Maggetti (2007, 2009), cannot be accommodated to the case of the Oversight Board. Instead, I use two idiosyncratic metrics: (1) Whether the Board concurred or overturned Meta's decision on a specific case and (2) the degree of burden Meta will incur by implementing the Board's recommendations. The first metric is a simple binary metric. The second is composed by mapping the unique Board's recommendations and sorting them into one of 10 types arranged in four tiers, each representing a greater degree of burden.

Seeing that available data indicates Meta implements most of the Board's recommendations, I assess which factors influenced Meta in its decisions: internal factors, relating to its relationship with the Board, or external factors, not directly connected to the Board. In evaluating external factors, I focus on two of them: the sentiment of Meta's press coverage and

Meta's stock price. To do so, I use a quantitative analysis of Meta's coverage in four major news outlets from 2018 to 2021, using the text analysis tool LIWC-22 to determine whether the coverage had a positive or negative sentiment, and various regression models to assess the difference in sentiment in 2021 (the first full year of the Board's operation) compared to previous years and possible connection between sentiment and stock price to Meta's reaction to the Board's recommendations. Internal factors are assessed using a qualitative method, based on interviews conducted with Board personnel, media interviews with Board personnel and public documents. I identify self-image themes and strategies the Board employs to enhance its autonomy and then correlate these themes with ones used by interdependent organizations to establish mutual respect and autonomy.

3. Evaluating Autonomy

3.1 Formal Autonomy

I first assess the Oversight Board's formal and de facto autonomy vis-à-vis Meta. The most relevant literature for this task is the literature which deals with the political independence of independent regulatory agencies (IRAs). Political independence relates to the independence of an IRA from governments, parliaments and politicians, specifically the degree to which an agency's decisions are made according to their legal mandate and without direct political control (Hanretty & Koop, 2012). The Oversight Board, not a traditional government agency by any means, was not formed by a government but by a private corporation. However, the role of Meta in creating and funding the Board closely parallels that of a government creating a regulatory agency, and therefore political independence provides a useful framework.

Political independence can be measured through two main aspects: formal independence and de facto independence. Formal or de jure independence primarily reflects the intent of the regulatory agency's creator, and is the degree of independence that flows from the legal instruments used to form and govern it (Gilardi, 2005; Hanretty & Koop, 2012). A governmental agency with a high degree of formal independence indicates that its creator intended to establish an autonomous powerful institution, while a lower degree indicates the intention to create a more servile body (Gilardi, 2005). Informal or de facto independence represents the effective independence of a regulatory agency in its day-to-day operations (Maggetti, 2007). If formal independence is a reflection of the legal status of an agency, de facto independence is a representation of its insulation from politicians (Donadelli & van der Heijden, 2022). De facto independence is relevant not only vis-à-vis elected officials, but also as it relates to the sectors or entities being regulated (Gilardi & Maggetti, 2011). This framework is especially relevant when examining the Oversight Board since Meta acts in the role of politician/government creating the agency, as well as that of the rule-taker the agency is tasked with regulating. I start with analyzing the Oversight Board's formal autonomy utilizing empirical tools created to assess the formal independence of regulatory agencies, which will enable to identify Meta's intention in creating it**.** For reasons discussed below, current empirical tools are not suitable to assess the Board's de facto autonomy, and that assessment is made using idiosyncratic metrics.

An early attempt to codify a regulatory agency formal or legal independence was made by Cukierman et al. (1992), who developed an index to assess the legal independence of central banks. This index uses four key indicators: appointment, dismissal and term office of the bank's governor; policy formulation cluster; objectives of the central bank; and limitations on the ability of the central bank to lend to the public sector (Cukierman et al., 1992). This index, however, is unsuitable to assess the autonomy of the Oversight Board, as most of its variables are specific to central banks and are not easily adaptable to other entities (i.e., who formulates monetary policy, objectives vis-à-vis price stability or limitations on lending to the government). Another index deals with the political independence of telecom regulators, and is also unsuitable to the case of the Oversight Board as it includes industry unique variables such as call rates or total number of telephone lines (Edwards & Waverman, 2006).

A more relevant sector specific index is Smithey & Ishiyama's (2000) judicial independence index, which assess the political independence of constitutional courts. The index offers six variables: 1) Can the court's decisions be overturned by other actors? 2) Does the court have priori judicial review? 3) Term length of judges, 4) Number of actors involved in nomination and confirmation, 5) Who sets the rules which determine the proceedings? And 6) The degree of difficulty in removing judges from office. Each variable is rated between 0 and 1, and with most variables the choice is binary (Smithey & Ishiyama, 2000). Though the Board is not a court by any measure, its role as an arbiter on specific content moderation cases lands a judicial aspect to its operation. However, this index is not fine-tuned enough and lacks variables that incorporate broader aspects of the Board's operation, such as the necessary distinction between the Board's co-chairs and members or its relationship with Meta.

The index developed by Gilardi (2002, 2005) is both more nuanced and generalized than previous indices and offers aspects that are not present in Smithey & Ishiyama's index. A modified and expanded version of Cukierman et al.'s index, Gilardi's index breaks down various aspects of a regulatory agency's operation into five primary indicators: status of the agency head, status of the members of the management board, relationships with government and parliament, financial and organizational autonomy, and regulatory competencies. These are sub-divided into simple quantifiable variables which are assessed and weighted on a sliding scale of 0 to 1, where a higher score indicates a higher degree of formal independence. To construct the index, each indicator is aggregated in the variable level by calculating the mean of each indicator. In the second step, the mean of the five indicator is calculated to produce the index's score. Each indicator has equal weight in the final score (Gilardi, 2002, 2005).

Gilardi's index, as well as the other indices, were criticized by Hanretty & Koop (2012; Koop & Hanretty, 2018), claiming they conflate bredth of powers with independence, when in essence the two are distinct from each other: bredth of powers represents the tools and activites under the agency's purview, whereas formal independence is the legal ability of an regulatory agency to make decisions without political interferance. "An agency may possess limited powers but exercise them independently; or it may possess a wide range of powers and exercise them with no independence" (Hanretty & Koop, 2012, p. 202). The authors also challenge assumptions inhernt in previous indices, such as absence of provision in the law with regards to term length and dismissal as indicating lower independence or scoring appoitnment by legislature higher than appointment by the executive, claiming local factors can render these assumptions moot or wrong. Other criticism focus on arbitrary weighting of items and assumed interval level of item responses. To counter these problems the Hanretty & Koop suggest to mesure independence and accountablity speprately, using two different indices. They also use a latent trait model based on item response theory (IRT) to score an regulatory agency independence or accountability, that can differentiate factors by weight (Hanretty & Koop, 2012; Koop & Hanretty, 2018).

To assess the Oversight Board's autonomy, I use augmented versions of Gilardi's and Hanretty and Koop's indices. Though these indices were developed for agencies created by governments, they can apply to assess the Oversight Board's autonomy with some adaptations. Most notably, questions regarding relationship with government and/or parliament were amended so that, where applicable, Meta takes the place of government, and the parliament option was omitted. As the Oversight Board's creator, Meta's role closely parallels that of the executive, whereas the company operates without an equivalent of a separate legislature.

Other changes were made to accommodate various idiosyncrasies of the Board's operation. Most notably, the internal organization of the board differs from that of other more traditional regulatory agency. While in entities like the Federal Trade Commission (FTC) or the Federal Reserve the commissioners are responsible for both regulatory decision affecting rule-takers and internal decisions such as budgeting and staff, in the case of the Oversight Board that role is split between the Board's trustees, which acts as one Board personnel describes, as "corporate managers", in charge of budget approval and hiring the Board's director and are considered by the staff the ultimate bosses; and between the Board members which make all content moderation decisions and policy advisories but are not normally involved in day-to-day operations (with the exception of the Board's four co-chairs which are involved in decision making on issues such as staffing and internal organization). The answers for relevant questions were adjusted to accommodate for these idiosyncrasies. Since the focus of this research is the decision making and policy impact of the Board's operation, the adjusted index does not directly assess the aspect of the Oversight Board's trust in its autonomy. However, where relevant answers were adapted to accommodate for the trust involvement in decision-making. These answers were scored lower compared to answers where Board members/co-chairs have full control, as the trust can be viewed as an element impeding the Board's decision-making process, at least partly because the trustees are appointed by Meta.

Finally, both Gilardi's and Hanretty & Koop's indices were created as a tool to enable quantative comparision between regulatory agnecies. The uniqness of the Oversight Board and the neccesary changes made to the index render this sort of comperassion moot in the current case. I therfore assess the Board's formal independence on a fuzzy low-to-high scale, based on a score of 0 to 1, where 0 represents no autonomy and 1 represetns full autonomy. After calculating the final score, I assign it a literal description based on the following key: 0-0.2 – low (level of formal autonomy); 0.21-0.4– low-medium; 0.41-0.6 – medium; 0.61-0.8 – medium-high; 0.81-1 – high. Answers were determend based on a questionnaire filled by a Board staff member, and verified through interviews and correspodance with staff and analysis of the Board's charter and bylaws.[1]

The adjusted Gilardi index assigned the Board a score of 0.6, indicating a medium level of formal autonomy verging on medium-high. I also scored an expanded and adjusted version of Gilardi's index, incorporating four additional variables suggested by Hanretty & Koop. In this

[1] See appendix I for the modified indices' questionnaire and answers.

version, the Board scored 0.64, indicating at medium-high level of formal autonomy. These alone, indicates that Meta intended to create an entity with considerablem yet somewhat limited, autonomy. However, though these scores are useful, they still provide only a partial view of the Oversight Board's autonomy, due to idiosyncrasies they cannot easily accommodate. Discussing these idiosyncrasies will provide a fuller appreciation of the Board's autonomy.

One important aspect is the Board's ability to revise the bylaws governing its operation. Normally, rules governing the operation of a regulatory agency are created by the government or parliament prior to its inception, sometime through legislation. Their revision is the sole prerogative of the state. A traditional agency cannot set rules governing its operation. The Oversight Board, however can revise its bylaws and indeed have done so in the past (Oversight Board, 2021b). Control over its bylaws enhances the Board's autonomy, as it allows it to shape its operational framework and adjust its commitments and scope of operation without necessitating an approval from Meta, a unique power not available to traditional regulatory agencies.

Another aspect is the status of Board members as part-time contract workers who maintain separate careers in the private market, academia or third sector. Traditionally the commissioners of a regulatory agency will be full-time employees committed to working exclusively for the agency. The members of the Oversight Board, however, all maintain separate and prominent careers and various fields in conjunction with their work for the Board. These include a senior editors at national newspapers, executive director of a NGO, several researches in well renowned academic institutions, prominent politicians including a former prime minister and one Nobel peace prize laureate (*Meet the Board | Oversight Board*, n.d.). This setup diminishes the Board members' dependency on the Oversight Board, and by proxy on Meta, and enables them greater autonomy in the decision-making process.

These idiosyncrasies enhance the Board’s autonomy in ways not captured by the indices, and lead to a conclusion that Meta created the Oversight Board with a high degree of formal autonomy, intending to design an entity that can develop into a reliable critical voice. However, this conclusion comes with significant caveat: The assessment was made by placing Meta in a role usually employed by government and/or parliament, and more specifically those institutions of a democratic regime for which the indices were developed. Meta, however, is not a government/parliament of a democratic country but a publicly traded corporation with one person in control of most voting shares – Founder and CEO Mark Zuckerberg. As such, it lacks the normal checks and balances that typify a democratic system and is more akin to a dictatorship in which the ruler can make binding decision at will. Though it created the Board with a high degree of autonomy, there is no system in place to make sure Meta maintains and respect its autonomy, as would be under a democratic administration. Meta can decide, according to considerations it is not obligated to share, to ignore the Board’s decision and recommendations, or to cut future funding, leaving the Board with no meaningful recourse and rendering the question of autonomy moot. Though the Oversight Board was created with a high degree of formal autonomy, this autonomy is conditioned on Meta’s good will.

3.2 De Facto Autonomy

Assessing the Oversight Board's de facto autonomy presents a more complicated challenge. Maggetti (2007) suggests an index which borrows from Gilardi's formal independence index, and adds metrics such as the agency's age, veto players, coordination of the economy, sectoral path dependence, and the effect of agencies' networks, to assess the agency's de facto independence (Gilardi & Maggetti, 2011; Maggetti, 2007). However, most of the added indicators are irrelevant to the Oversight Board, and the index cannot be adapted to accomdate its unique situation. Specifically, indicators such as degree of coordination between national economies in western countries or mode of regulation used before the creation have no relavnce when assessing the Oversight Board and cannot be modified without changing the initial meaning. Creating a wholly new index is currenlty unwaretnend, due to the lack of comparable entities to which said index could be applied, and is outside the scope of the current research.

Instead, to assess the Oversight Board's de facto- or behavioral-autonomy, I use two idiosyncratic metrics created using the Board's decisions: one based on the Board's binding judgements and the other on the Board's non-binding recommendations. Judgements are assesed on the basis of wheater the Board sided or contradicted Meta's decision; as it is evidanet that dissenting judgmets represent greater autonomy and advarsarial capacity. Recommendations are assesd by the degree of burden their implementaion represents. Implementaton of a more onerous recommendation requiares more resources – time, money, personal – or incurs a loss of autonomy for Meta, indicating a larger degree of de facto autonomy by the Board.

The Board's judgments are analyzed using a simple binary metric: whether the Board upheld or overturned Meta's final decision in the case, as generally this was the standing decision during the Board's deliberations. Of 23 cases selected by the Board and finalized between the 28th of January 2021 and the 1st of February 2022, the Board has made judgment in 22 of them.[2] Of those, the Board has overturned Meta's final decision in 11 cases, a number representing moderate degree of de facto autonomy. This number differs from the Board's own analysis, as presented in its yearly report, which uses as a reference point Meta's original decision rather than its final one (Oversight Board, 2022b).

However, of the 11 cases where the Board upheld Meta's final decision, in 6 cases Meta has overturned its original decision after the Board selected the case for review; in some cases, after the user has gone through numerous appeal process to no avail. In effect, the Board's process resulted in overturing 17 of Meta's original decisions. The fact that in more than a quarter of the cases the mere selection of a case by the Board was enough to trigger an internal review process in which Meta voluntarily overturned its decision, and that in most of the remaining cases the Board has overturned the final decision, indicates a high level of de facto autonomy.

The second metric is based on the Oversight Board's policy advisory statements (or recommendations). These represents the Board's broader view of Meta's operation, beyond the scope of a specific case, and are often the forefront of its criticism. Unlike the case decisions, the recommendations are non-binding, though Meta is obligated to reply within a specific time frame. They represent the Board's wider philosophy and their coverage in the media can be a source in further strife and pressure for Meta. Scoring a recommendation by the burden its implementation will incur is a good proxy for assessing the Board's de facto autonomy: recommendations which implementation will require more resources or will diminish Meta's autonomy represent a greater willingness by the Board to act in an adversarial critical role, in-contrast to appeasement and

[2] In one case, the post was deleted by user before the Board could review it.

serving as an instrument to deflect criticism, and therefor indicate a high degree of de facto autonomy.

To assess this, the Board's statements were first broken down into 95 unique recommendations. This number does not correspond to the number of recommendations presented by the Board (Oversight Board, 2022b) for two reasons. First, what the Board considers as a single recommendation might include two or more distinct recommendations. For example, the following single statement consists of 3 recommendations: "(1) Inform users when automated enforcement is used to moderate their content, (2) ensure that users can appeal automated decisions to a human being in certain cases, and (3) improve automated detection of images with text-overlay so that posts raising awareness of breast cancer symptoms are not wrongly flagged for review" (*Case Decision 2020-004-IG-UA*, 2021). Second, a few recommendations are repeated in different decisions, and were eliminated from the count.

The 95 recommendations were than mapped into ten distinct types reflecting the nature of the recommendation (policy change, operational change, etc.). These are grouped into one of four tiers based on the burden they lay on Meta vis-à-vis resource allocation and/or the degree of autonomy the company will lose by implementing them. A higher tier corresponds to a higher burden or greater loss of autonomy. Some recommendations fit into more than one type or into more than one tier. In scoring those, only the higher tier is considered. Types' placement inside tiers is random and non-hierarchical.

Tier 1: implementing the recommendation will require minimal resources and does not affect Meta's policy or autonomy.

Type 1: Case specific – minor recommendation regarding discussed case only that requires minimal resources. Ex.: “[the decision will] only be implemented pending user notification and consent” (*Case Decision 2020-007-FB-FBR*, 2021).

Type 2: Minor adjustment – Meta is asked to correct a simple error or make minimalistic changes to policy and/or procedure. Ex.: “restore the misplaced 2017 guidance to the internal guidance for content moderators” (*Case Decision 2021-006-IG-UA*, 2021).

Tier 2: implementing the recommendation will require non-nominal resources and/or enable greater scrutiny of Meta’s operations; minimal effect on autonomy.

Type 3: Operational change – Meta is asked to implement new procedure, mostly vis-à-vis individual user communication. It does not affect policy or decision making, but mostly the way they are communicated to users. Implementation might require development of new tool or tech, but these are relatively simple and are not the main goal of the recommendation. Ex.: “Ensure that users are always notified of the reasons for any enforcement of the Community Standards against them, including the specific rule Facebook is enforcing” (*Case Decision 2020-003-FB-UA,* 2021).

Type 4: Transparency change – Instructs Meta to reveal data it already has or can generate easily, or to clarify an obscure policy. Require minimal resources but might enable greater scrutiny and criticism of the company. Ex.: “Explain and provide examples of the application of key terms from the Dangerous Individuals and Organizations policy” (*Case Decision 2020-005-FB-UA*, 2021).

Tier 3: implementing the recommendation will require significant resources and/or slightly diminish Meta’s autonomy.

Type 5: Develop policy and/or procedure – Meta is asked to develop or evaluate a certain policy or implement a new procedure, but the Oversight Board doesn't dictate or shape the policy/procedure beyond a few general guidelines. Ex.: "develop and publish a policy that governs its response to crises or novel situations where its regular processes would not prevent or avoid imminent harm" (*Case Decision 2021-001-FB-FBR*, 2021).

Type 6: Conduct a report/study/audit – Meta is asked to study and report on its operation, one-time or periodically. This recommendation requires Meta to create new data and demands dedicated resources; the Oversight Board might stipulate data to be included or set guidelines. Ex.: "Conduct a reviewer accuracy assessment; Study the impacts on reviewer accuracy when content moderators are informed they are engaged in secondary review" (*Case Decision 2021-012-FB-UA*, 2021).

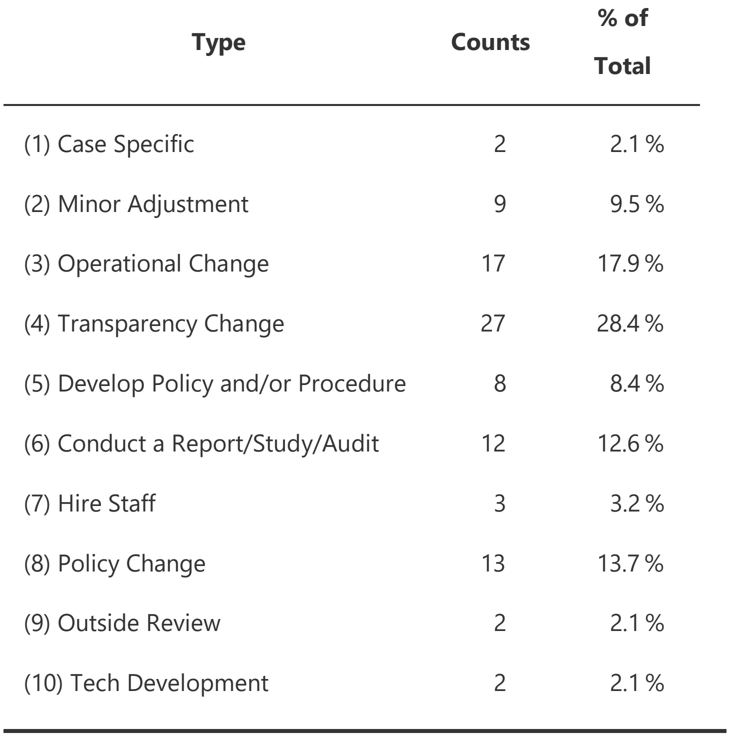
Type 7: Hire staff – Meta is asked to hire staff, in order to achieve a specified goal or to implement new procedure. Ex.: "Restore both human review of content moderation decisions and access to a human appeals process to pre-pandemic levels" (*Case Decision 2021-003-FB-UA*, 2021).

Tier 4: Implementing the recommendation will require very significant resources and/or will considerably diminish Meta's autonomy

Type 8: Policy change – Meta is asked to make changes to policy as dictated by the Board. In contrast to type 5, here the Oversight Board stipulates how the policy should look, which represents a greater impact to autonomy. Ex.: "Revise Instagram's Community Guidelines to specify that female nipples can be shown to raise breast cancer" (*Case Decision 2020-004-IG-UA*, 2021).

Type 9: Outside review – Meta is asked to allow a third-party access to internal system and/or staff to create a report on its operation. Meta cannot control the result. Ex.: "engage an independent entity not associated with either side of the Israeli-Palestinian conflict to conduct a thorough examination to determine whether Facebook's content moderation in Arabic and Hebrew have been applied without bias" (*Case Decision 2021-009-FB-UA*, 2021).

Type 10: Tech Development – Instructs Meta to develop new technology and/or tools. Implementation can require considerable resources. Ex.: "improve automated detection of images with text-overlay" (*Case Decision 2020-004-IG-UA*, 2021).



Type	Counts	% of Total
(1) Case Specific	2	2.1 %
(2) Minor Adjustment	9	9.5 %
(3) Operational Change	17	17.9 %
(4) Transparency Change	27	28.4 %
(5) Develop Policy and/or Procedure	8	8.4 %
(6) Conduct a Report/Study/Audit	12	12.6 %
(7) Hire Staff	3	3.2 %
(8) Policy Change	13	13.7 %
(9) Outside Review	2	2.1 %
(10) Tech Development	2	2.1 %

Table 1: Oversight Board recommendations divided by type

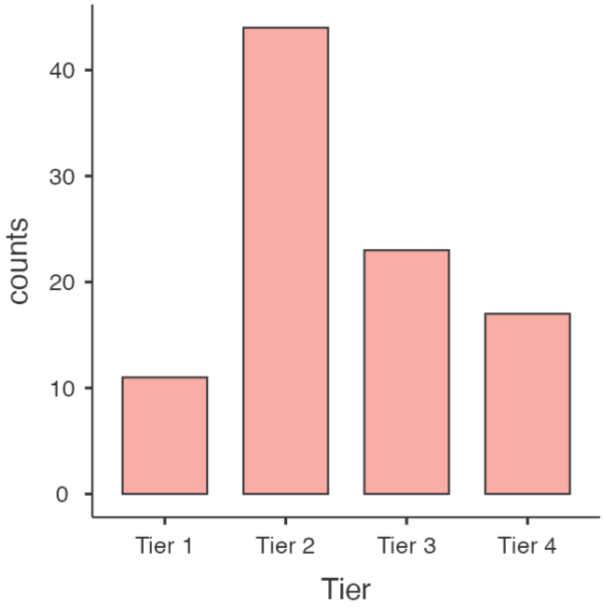


Figure 1: Oversight Board recommendations divided by tier

Most recommendations fall into one of the two lowest tiers, with almost half in tier 2. These represent recommendations which implementation will be relatively simple, and negligent in terms of resources. On their own, they do not represent a meaningful burden on Meta. However, these recommendations also have a substantial culminative effect. If Meta where to adopt them all, they will represent a considerable change to the company's daily operations. The 17 type 3 recommendations will transform users access to appeals options and provide relevant and useful knowledge about actions taken against them. Implementation of the 28 type 4 recommendations will have a broad effect on the understanding of Meta's content moderation policy and enforcement efforts, enabling a more significant scrutiny of its operation and potentially opening the company to broader criticism. Not less important is the underlaying philosophy of the tier 2 recommendations, which represent a consistent approach: greater transparency and clarity in policy presentation and execution. If Meta were to apply the logic behind these recommendations to other aspects not considered by the Board, it will transform its content moderation practices to operate with greater transparency, allowing users and outside observers to understand the company’s decision making and execution, changing a once obscure process.

Though they are only the minor share of recommendations, the more demanding tier 3 and tier 4 recommendations are still numerous and represent a considerable burden. In tier 3, most notable are the 8 type 5 and 12 type 6 recommendations, which implementation will require considerable resources to study or collect data and to assemble reports or develop policies. Implementing the type 5 recommendations will change or for the first time formalize the way Meta deal with content such as satire, content relating to dangerous individuals and organizations or bullying, or the way the company addresses novel situation – all of them represent major aspects of content shared on Meta’s platforms. In tier 4, there are 13 type 8 recommendations. These dictate specific changes or policies Meta should adopt in areas such as regulated goods, safety in online speech or the influential user policy, and shapes the topics and themes users are allowed to converse about in Facebook and Instagram. Viewed together, these recommendations can considerably affect the conversation in Meta’s platform and the way the company interacts with these users.

In analyzing the Board’s decisions, it is apparent that the Board is willing the act in an adversarial capacity, creating recommendations that challenge Meta and which implementation will produce significant change in the company’s operations and policies and will place it in under greater scrutiny. This analysis indicates that the Oversight Board operates with a high degree of de facto autonomy and can create recommendations that challenge Meta and provides severe critic of the company’s policies and operations.

However, the ability of the Oversight Board to operate autonomously and fill an adversarial role is only significant if Meta adopts and implements the Board's recommendations. I now turn to examine Meta's responses to the Board's recommendations, and the factors influencing them.

4. Evaluating Influence

The final aspect of this analysis is an evaluation of the impact the Oversight Board had on Meta. In other words, was the Board successful in changing Meta's content moderation policies and promoting its vision for content moderation and what are the reasons for its success or failure. Evaluating the Board's influence on Meta is a complicated task which first and foremost requires more time. Implementation of policy and operational changes is a long-term endeavor and meaningful results may not be observable for some time. This sort of evaluation is not possible under the time scope of this paper and would have to await further research. However, it is possible to observe Meta's pro forma willingness to accept the Oversight Board's policy recommendations and the first steps taken to implement them through Meta's responses to recommendations and the Board's monitoring of their initial implementation efforts.

This data can be obtained from two publicly available source: Meta's Transparency Center's web page detailing its responses to the Oversight Board's recommendations, and the Board's annual transparency report, analyzing Meta's implementation of said recommendations. According to Meta's own data, reviewed on the 20th of November 2022, of the 119 responses it made public between the 25th of February 2021 and the 16th of August 2022, it outright rejected only 15.1% of them. The largest share, 33.6%, represents recommendations Meta announced it will implement fully, with a further 26.9% it means to partially implement. 12.6% of recommendations Meta categorized as Assessing Feasibility, in essence accepting the Board view but requiring further checks as to their practical implementation. The remaining recommendations were categorized as Work Meta Already Does, meaning the recommendation was implemented prior to the Board making it (Meta, n.d.).

In its first annual report, released June 2022, the Oversight Board tracked whether Meta implemented its recommendations, and to what degree. Of the 86 recommendations covered in the report[3], the Board independently confirmed, utilizing data provided by Meta and a so-called "data-driven approach", that Meta fully implemented 16.3% of them. In addition, Meta reported progress with 48.8% of the recommendations, and a further 24.4% of recommendations were categorized as implemented without the Board able to verify Meta's claim, or as work Meta already does. The remaining 10.5% are recommendations Meta rejected (Oversight Board, 2022b). The Board updated these figures in August 2022 and again in October 2022, as part of its quarterly transparency reports. According to the October 2022 update, part of the 2nd quarter transparency report, out of 118, the Board verified that Meta fully implemented 17.8% and partially implemented 3.4%. 22.9% were reported by Meta as fully implemented or as work it already does and on 36.4% Meta reported progress – both could not be verified by the Board. Meta declined to implement 6.8% of recommendations after completing a feasibility assessment and omitted, declined or reframed a further 12.7%. In total, Meta accepted and implemented fully or partially 80.5% of the Board's recommendation – a very high acceptance rate. Recommendations implemented include some that were or would be classified as high tier recommendations, such as codifying policy response to global crises or releasing a report conducted by an independent third party assessing Meta's policies in Israel and Palestine; as well as recommendations with direct impact to users, such as increased granularity of user notifications events like government takedown requests or content moderation policy violations (Oversight Board, 2022e).

[3] The Board's count differs from this paper's count, with some repeating recommendations counted twice by the Board, while other multi-layered recommendations split into two or more by the author. See p. 31 for a more detailed explanation.

The above data suggests that the Oversight Board was demonstrably successful in influencing Meta's policy and affecting changes to its operation, in those areas where it set out to do so. This than leads to the question, how can the Board's success be explained? The rest of this paper will strive to answer it. This will be done by examining two central hypotheses:

H1: The Oversight Board's success can be explained by Meta responding to actions taken by third parties, such as media and regulators, pressuring it to change content moderation policies.

H2: The Oversight Board's success can be explained by Meta responding to actions taken by the Board itself, establishing its respect and autonomy vis-à-vis Meta.

4.1 External Factors

The most relevant research to help assess the first hypothesis concerns the reasons corporations engage in socially responsible behavior. A common measure of whether a corporation acts in a socially responsible way examines (1) if a corporation does not knowingly acts in a way that harms its stakeholders (such as users, employees, investors and suppliers) and (2) if once causing harm it acts to rectify it once discovered (J. L. Campbell, 2007). Meta's responses to the Oversight Board's recommendations can be viewed as an act of social responsibility: The Board has identified areas where Meta's operations caused harm to a group of stakeholders (i.e., its users), and suggested way to mitigate it. Meta's responses to these recommendations are therefore a decision whether to act or not to act in a socially responsible way.

Research suggests several ways in which corporations might be influenced to display a more socially responsible behavior. One proposition is that corporations will act in socially responsible ways when independent non-government organizations, such as the press, monitor and criticize their behavior (J. L. Campbell, 2007). The press plays an increasingly significant role in corporate governance, with news organizations using the tools of public exposure to critic corporations and pressure them to change behavior and policies. This pressure has led corporations to dedicate considerable resources to media relations (J. L. Campbell, 2007; Kjær & Langer, 2005). This has been especially notable in Meta's case, where reporting by news organization has repeatedly revealed problematic behavior by the company. Reports such as the expose by The Guardian and The New York Times of the "Facebook–Cambridge Analytica data scandal", which revealed how a British data analytics company harvested data of 87 million Facebook users and used it to create psychological profiles in order to display targeted political ads, created and international outcry, which included congressional hearings and regulatory investigations, leading Meta to change its policies around access to user information and paying a fine of $5 billion in the US (Cadwalladr & Graham-Harrison, 2018; Egan & Beringer, 2018; *Facebook, Social Media Privacy, and the Use and Abuse of Data | United States Senate Committee on the Judiciary*, 2018; Rosenberg et al., 2018; Schroepfer, 2018; J. C. Wong, 2019).

Another factor influencing corporations' tendency to act in socially responsible ways is their financial performance. Corporations act to maximize profit and value to shareholders, and research suggests that firms with weak financial performance are less likely to act in a socially responsible way. This, due to lesser availability of resources to fund socially responsible behavior, when to compared to profitable corporations (J. L. Campbell, 2007; Margolis & Walsh, 2001; Orlitzky et al., 2003; Waddock & Graves, 1997). Other research concluded that institutional investors would review a company's social performance when considering whether to acquire or dispose of a company's stock and would invest more in corporation with better social performance. Firms that allocate resources towards socially responsible activities do not incur detrimental impact or penalty. A company suffering from poor stock performance might therefor seek in boost its stock price and attractiveness to investors by engaging in socially responsible behavior (Coffey & Fryxell, 1991; Graves & Waddock, 1994; Mahoney & Roberts, 2007; Teoh & Shiu, 1990).

Press coverage and stock price can also serve as a proxy to other factors that influence corporates' social behavior, such as congressional hearing and legislations from politicians or regulatory actions. These acts will often be reported by the press, especially true with regards to Meta which tends to generate considerable press coverage, and might affect stock price negatively (for example, institutional investors, fearing a regulatory action will hurt a corporation financial performance or competitiveness, might decide to divest their holdings). I analyze the possible influence of these two external factors – press and stock price – on Meta's willingness to accept the Board's recommendations, through a quantitative examination of the sentiment, or tone, of Meta's press coverage and its performance in the stock market. This examination will enable to assess the influence of these two main factors, and indirectly the influence of other factors.

To assess the influence of press coverage I first collected all the articles concerning Meta and published between 2018 and 2021 in The New York Times (NYT), The Wall Street Journal (WSJ), The Associated Press (AP) and Financial Times (FT). 2021 was chosen has the reference year being the first full year of the Oversight Board operations, with its first decisions and recommendations presented in January. A more negative tone in press coverage in 2021 compared to the immediately preceding years might indicate that press coverage of Meta was a factor in its decision to accept the Board's recommendations. NYT and WSJ were as chosen each of them is regarded a "paper of record" with a wide distribution in the US and other countries, and considerable influence on lawmaker, regulators and the news cycle. AP is one of the largest news agencies in the world, and its stories are syndicated to thousands of news outlets. FT is the largest financial newspaper in Europe and is widely read among decision makers in the European Union's member countries and in the European Commission, the executive branch of the EU.

The news articles were searched and downloaded on a yearly basis from Nexis Uni for NYT, AP and FT and from ProQuest Central for WSJ, using search queries for the appearance of either of the following terms in headline or sub-headline: "facebook", "meta"[4], "instagram", "whatsapp", and "zuckerberg". Though some articles concerning Meta might not include these terms in the headlines (such as ones reporting on a few large tech companies in aggregate), they strike a good balance between collecting as many relevant stories as possible and avoiding too much unrelated stories.

In total, 5,730 articles were downloaded. These were than manually scanned to identify and remove duplicates, texts with 100 words or less and articles not mainly concerning Meta, its operations or its users. The latter included aggregated articles or newsletters summing various news events, and articles using one of Meta's brands as "click-bait" (for example, a WSJ article about the real-estate market in Tuscany, using the term "Instagram worthy" in the headline while text itself included only a passing mention of the social network, was removed). The elimination process resulted in a dataset of 4,030 articles, according to the division presented in table 2.

[4] The "Meta Platforms" moniker, commonly shortened to "Meta", was introduced in October 2021, but was used in all searched years for the sake of continuity.

```
       year |      Freq.     Percent        Cum.
------------+-----------------------------------
       2018 |      1,432       35.53       35.53
       2019 |        996       24.71       60.25
       2020 |        779       19.33       79.58
       2021 |        823       20.42      100.00
------------+-----------------------------------
      Total |      4,030      100.00
```

Table 2: Number of articles about Meta in NYT, FT, AP and WSJ divided by year

To analyze the sentiment of coverage I used the text analysis tool LIWC-22 (Linguistic Inquiry and Word Count), which can identify various linguistic traits in texts (Tausczik & Pennebaker, 2010). LIWC-22 uses two baseline metrics to assess the sentiment of a text: precent of positive words in the text (such as “good”, “well”, “new”, “love”), and precent of negative words in the text (such as “bad”, “wrong”, “too much”, “hate”). These two are than combined into a single “Tone” matric, which scores the degree of positive/negative sentiment of each text on a scale of 1 through 100, with a higher score indicating a more positive sentiment. A score of less than 50 is an indication of a negative sentiment (Boyd et al., 2022).

The descriptive statistic results for the dataset, with the means for two baseline metrics – percent of positive words (tone_pos) and percent of negative words (tone_neg) – and the composite metric (Tone) categorized by year, are presented in Table 3.

```
    year |  tone_pos  tone_neg      Tone
---------+------------------------------
    2018 |  1.713233  1.509888  26.15733
    2019 |   1.79012  1.388002  28.23991
    2020 |  1.787279  1.582092  25.36718
    2021 |   1.73158   1.58305  24.84089
---------+------------------------------
   Total |  1.750295  1.508663  26.25045
----------------------------------------
```

Table 3: the mean of percent of positive words (tone_pos), negative word (tone_neg) and overall sentiment (Tone) in articles about Meta, divided by year

It is immediately apparent that while the sentiment for 2021 was more negative than any of the preceding years, Meta's coverage in general tends heavily towards a more negative sentiment. This sentiment is an outlier compared to general news coverage. The developers of LIWC-22 established a baseline of sentiment coverage in the NYT by randomly selecting and analyzing 1,000 texts with 100 words or more. The mean for the composite Tone metric was 37.08 (Boyd et al., 2022). In comparison, the mean for NYT's articles in the dataset is 28.63.

This research interest is in whether the sentiment for 2021 differed significantly from previous years. To determine that, a simple linear regression model was used to test if the year of coverage, with 2021 as the baseline, significantly predicted the sentiment of coverage, as represented by the composite Tone metric. Since the data does not confirm to the heteroskedasticity assumption, the model was used with robust. The dataset fits a Poisson distribution, but linear model was chosen as the large number of observations negates the need to adjust for the violation of the normality assumptions. However, the significance of coefficients in a Poisson regression model with robust was almost identical to the linear regression results.[5] All other linear regression assumptions hold. Results of the linear regression model are presented in table 4.

[5] See appendix II for Poisson regression results.

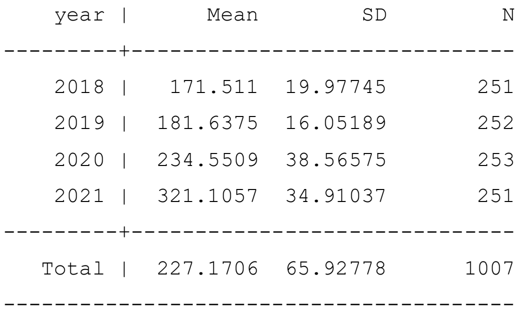
```
Linear regression                               Number of obs     =      4,030
                                                F(3, 4026)        =       8.64
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0067
                                                Root MSE          =     15.086

------------------------------------------------------------------------------
             |               Robust
        Tone | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        year |
       2018  |   1.316438   .6509787     2.02   0.043     .0401599    2.592717
       2019  |   3.399023   .7216259     4.71   0.000     1.984237    4.813809
       2020  |   .5262889   .7358081     0.72   0.474    -.9163023     1.96888
             |
       _cons |   24.84089   .5178097    47.97   0.000     23.82569    25.85608
------------------------------------------------------------------------------
```

Table 4: Linear regression model for Tone of Meta’s coverage with year as predictor

The fitted regression model was Tone = 24.8409 + 1.3164*(“2018”) + 3.399*(“2019”) + 0.5263*(“2020”). The overall regression was statistically significant (R^2=0.0067, F(3, 4026)=8.64 P<0.001). The sentiment for 2018 coverage was significantly more positive compared to 2021 (P<0.05), as well as the sentiment for 2019 (P<0.001). There was no significance difference between 2020 and 2021 (P>0.05).

For the first full year of the Oversight Board’s operation, the sentiment around Meta’s coverage was not significantly more negative than the coverage in the immediately preceding year, though it was significantly more negative than the two years prior. It is notable that the coverage for 2021 was significantly more negative compared to 2018, the year in which Meta decided to create the Board – a possible indication of an external factor influencing Meta’s attitude towards the Board’s recommendations. However, this model only indicates that the sentiment of Meta’s coverage could have acted as an external factor. To establish a more direct link, different dataset and models were used, and will be presented later.

First, I examine the possible impact of Meta's stock price on its responses to the Oversight Board's recommendations. To assess this, I used the online archive of Yahoo! Finance to download the closing price of Meta's stock for each day of trading between 2018 and 2022, and then calculate the mean for each year. Results are presented in Table 5. The mean stock price for 2021 is higher than any of the preceding years. Since research connects lower stock price to a corporation decision to engage in socially responsible behavior, it is apparent that Meta's stock price cannot be considered an outside factor influencing Meta, whether significant or not, and no further analysis of this dataset is needed.



```
    year |      Mean        SD         N
---------+------------------------------
    2018 |   171.511  19.97745       251
    2019 |  181.6375  16.05189       252
    2020 |  234.5509  38.56575       253
    2021 |  321.1057  34.91037       251
---------+------------------------------
   Total |  227.1706  65.92778      1007
----------------------------------------
```

Table 5: Meta's mean stock price by year

To identify a possible direct connection between coverage sentiment and Meta's responses a different dataset was compiled and multinomial logistic regression and a logistic regression models were used. In this analysis, the dependent variable ("action") was Meta's responses to Oversight Board's recommendations, published between the 25th of February 2021 and the 16th August 2022, as obtained from company's transparency website (Meta, n.d.). The dependent variable has 119 observations, each a response to an Oversight Board recommendations, according to the division presented in table 6. One category, "Work Meta already does", denotes cases where the Board's recommends action Meta has taken prior and does not include a component of decision on Meta's part. Observations in the category were therefore dropped from all models.

```
               action |      Freq.     Percent        Cum.
----------------------------------+-----------------------
 Assessing feasibility |         15       12.61       12.61
    Implementing fully |         40       33.61       46.22
  Implementing in part |         32       26.89       73.11
     No further action |         18       15.13       88.24
Work Meta already does |         14       11.76      100.00
----------------------------------+-----------------------
                 Total |        119      100.00
```

Table 6: Meta’s responses the Oversight Board recommendations, divided by type or response

The explanatory variables were the average sentiment of Meta’s coverage in NYT, WSJ, AP and FT for the 30 days preceding the response (“tone_avg30day”), and Meta’s average stock price in the 30 days preceding the response (“stock_avg30day”). The data was obtained and compiled according to the methods of the two previous analyses. The 30 days period was selected as initially Meta was allotted 30 days to respond to the Board’s recommendations. The response period was extended to 60 days in February 2022 (Oversight Board, 2022a), but the 30 days analysis period was maintained for the sake of continuity. The controlling variables were the Oversight Board’s age in weeks on the day of Meta’s response (“board_age_weeks”), controlling for the possibility the Board gains more legitimacy the longer it operates, and the number of articles published in NYT, WSJ, AP and FT in the 30 days period (“article_num30day”), controlling for a period with outsized- or undersized-coverage.

The first model is a multinomial logistic regression with “No further action” as the base category, being the only cases where Meta outright refused implement, in full or in part, or even consider implementing, a Board recommendation. This model has 105 observations. The overall regression was statistically insignificant (Pseudo R^2=0.0734, P>0.05). None of the coefficients was found to be statisticaly significant, in either of the reference categories. Results are presented in table 7.

```
Multinomial logistic regression                        Number of obs =    105
                                                       LR chi2(12)   =  20.20
                                                       Prob > chi2   = 0.0634
Log likelihood = -127.46055                            Pseudo R2     = 0.0734

-------------------------------------------------------------------------------------
           action_num | Coefficient  Std. err.      z    P>|z|     [95% conf. interval[
----------------------------------------------------------------+--------------------
Assessing_feasibility|
        tone_avg30day |  -.0418812   .1632657    -0.26   0.798    -.3618762    .2781137
       stock_avg30day |  -.0048426   .0088755    -0.55   0.585    -.0222382     .012553
      board_age_weeks |   .0257246   .0266691     0.96   0.335    -.0265459    .0779951
     article_num30day |   .0124449   .0117156     1.06   0.288    -.0105172     .035407
_               cons |  -.1307284   4.844846    -0.03   0.978    -9.626452    9.364995
----------------------------------------------------------------+--------------------
Implementing_fully|
        tone_avg30day |  -.0819231   .1294073    -0.63   0.527    -.3355567    .1717104
       stock_avg30day |  -.0000739   .0064312    -0.01   0.991    -.0126788    .0125309
      board_age_weeks |   .0089873   .0195474     0.46   0.646    -.0293249    .0472996
     article_num30day |   .0062733   .0101891     0.62   0.538     -.013697    .0262436
_               cons |   1.908277   3.695719     0.52   0.606    -5.335199    9.151752
----------------------------------------------------------------+--------------------
Implementing_in_part|
        tone_avg30day |   .1597947   .1427066     1.12   0.263    -.1199051    .4394945
       stock_avg30day |  -.0035575    .006439    -0.55   0.581    -.0161778    .0090628
      board_age_weeks |  -.0273093   .0189807    -1.44   0.150    -.0645107    .0098921
     article_num30day |   .0084064   .0107861     0.78   0.436    -.0127339    .0295467
_               cons |  -1.530713   4.154872    -0.37   0.713    -9.674114    6.612687
----------------------------------------------------------------+--------------------
No_further_action     |  (base outcome(
-------------------------------------------------------------------------------------
```

Table 7: Multinominal logistic regression model with Meta's response to the Oversight Board's recommendations as dependent variable.

The second model was a logistic regression with a new dependent variable ("action_bin"). transformed from the previous dependent variable to include only two categories: "No further action", being the base category, and "Assessing feasibility", "Implementing fully" and "Implementing in part", combined into a single category representing all the cases where Meta agreed, at least partly or in principle, with the Board's recommendations. Here too the overall regression was statistically insignificant (Pseudo R^2=0.0142, P>0.05). None of the coefficients was found to be statistically significant. Results are presented in table 8. It is notable that, controlling for all other variables, the log odds for the average sentiment of coverage in the 30 days period was higher for cases where Meta agreed fully or partially with the Board's recommendations, compared to cases where it refused to implement them. This outcome can indicate that a more negative press coverage did not influence Meta to accept the Board's recommendations, negating the possibility of press coverage being an external factor influencing Meta. But, since the results are insignificant, one cannot make a firm assumption such as this.

```
Logistic regression                                     Number of obs =    105
                                                        LR chi2(4)    =   1.36
                                                        Prob > chi2   = 0.8507
Log likelihood = -47.42398                              Pseudo R2     = 0.0142

------------------------------------------------------------------------------------
      action_bin | Coefficient  Std. err.      z    P>|z|     [95% conf. interval[
-------------------------------------------------------------+------------------
   tone_avg30day |   .0210551    .118387     0.18   0.859    -.2109791    .2530893
  stock_avg30day |  -.0037074   .0057082    -0.65   0.516    -.0148952    .0074805
 board_age_weeks |  -.0043203   .0173892    -0.25   0.804    -.0384026    .0297619
article_num30day |   .0081972   .0096445     0.85   0.395    -.0107056    .0271001
           cons |   1.831341   3.420227     0.54   0.592    -4.872181    8.534862
_
------------------------------------------------------------------------------------
```

Table 8: Logistic regression model with Meta's response to the Oversight Board's recommendations as dependent variable.

A few factors imped the accuracy of these models. Mainly, the low number of overall observations (n=105), and the low number of observations in the reference category in relation to the other categories (17.1% of all observations). And though the results do not allow to establish the influence of external factors, they do not allow to preclude them. Further data, collected over a longer time, is needed to reach any concrete conclusion. Currently, however, the data does not allow for the acceptance of the first hypothesis.

4.2 Internal Factors

An assessment of the feasibility of the second hypothesis requires a qualitative approach, specifically one that analyzes the way the Oversight Board defined itself in relation to Meta, and the strategies it employed to actualize its self-image and establish its autonomy and respect vis-à-vis Meta. This is done through an in-depth overview Oversight Board's first year and a half of operation, based on interviews conducted with several Board personnel,[6] analysis of dozens of media interviews given by Board members and staff, and review of publicly available documents created by the Board and Meta. I first present two key themes in the Oversight Board's self-image and perceived role. I than map three practical strategies the Board uses to support and promote its self-image perception. Finally, I discuss how these strategies correspond with strategies used by interdependent organizations to establish autonomy and mutual respect.

The First self-image theme is the Oversight Board's presonnel conception of the Board as an autonomous not beholden to Meta's considerations. This is apperant in subtle ways, such as the Board's emphiss on publishing seliant policy recommendations and its belief that they will have an impact on Meta's policy and operation. It is also apperant directly when Board perosnnel discuss its autonomy. One interviewee said: "We are a like the Golem who has struck his creator, in their eyes (…) In our discussions we very much emphasize that it is quite possible that our policy decisions will place a very heavy burden on Facebook. But it's not relevent to us".

This theme is espacially seliant in media interviews. Board member Michael McConnell commented on this by referring to the importance of his and other Board members standing in a March 2022 interview: "[These are] people whose substantial reputations will enable them to

[6] "Board personnel" denotes either Board members or the Board's hired staff. Due to the small number of potential interviewees and to further preserve anonymity, there is no distinction between Board members and staff.

decide these cases the way they think. They’re not going to be deciding them the way Facebook wants. And they're not going to decide them the way Twitter storms may want” (Lloyd, 2022).

Board member Emi Palmor raised a similar issue in a May 2020 media interview:[7] “The people in the committee are not weaklings, they all have status, positions, and a public image to maintain and they are not going to throw all that away on Facebook’s behalf. We are a group of people that are getting into this rather ambivalently due to the severity of the incidents Facebook was involved with in the past but we intend to use this regulatory startup to make a difference” (Kabir, 2020a).

Other members stressed the importance of the charter and bylaws in establishing the Board’s autonomy. “The Board’s governing instruments (the Charter and Bylaws) establish institutional, functional, budgetary, and personal guarantees that members can act with complete autonomy from the economic, political, or reputational interests of the company”, said Board co-chair Catalina Botero Marino in a March 2021 media interview. “For example, we operate with a non-revocable endowment of US$130 million that is administered by a trust independent of the company; our appointment is for a fixed term that the company cannot interrupt; the Board's administration is completely independent; and members do not depend in any way on the company. Beyond these institutional elements, the decisions that we have taken thus far – the majority of which overturn those made by Meta – clearly show that we are not shy about holding it to account” (Marino & Tuchtfeld, 2021).

Board member Afia Asantewaa Asare-Kyei reflected on this theme from a more personal viewpoint in a May 2020 media interview: “I have no intention or interest in protecting Facebook and I would not have accepted this role if I believed that the board could be used as a shield for

[7] The media interview was conducted by the author of this paper.

Facebook. The Board does not take responsibility away from Facebook, it introduces a new level of oversight that will make Facebook more accountable and improve the way they make decisions. The Oversight Board will hold Facebook to account, and will both scrutinize and publish how the company implements binding decisions and policy recommendations" (Atoklo, 2020).

This theme is also significant for the Board's staff, as is evident in a May 2021 interview with Dexter Hunter-Torricke, lead communicator for the Oversight Board: "The board is really structured in a way to keep Facebook at arm's length as much as possible and entirely on case decisions. [None of the members] has any interest in defending Facebook or supporting their interests. And lots of them have been very critical about Facebook, which I think is a great sign of the independence of the board as well" (Arenstein, 2021).

Indeed, the independence and autonomy of the Board remains a consistent throughline in interviews and discussion throughout its operation.

The Sencond self-image theme is the precived role the Board has in the eyes of its members and staff. A Board perosnnel interviewed believes the Board's main role is to act as an idea accelarator for high-level content moderation policy: "[Our role is to] be one step ahead, even ten steps ahead. At least at the ideas level. This is our privilege (…) We don't need to develop the product, we don't need to update the business model". And later: "We see ourselves as developing something very, very pioneering in self-regulation". This role, the interviewee believes, facilataed substantial changes in Meta, which acording to them started developing new tools and policies in respose to the Board's recommendations; though they did state it is still too early in the process to make a true evaluation the Board's impact.

Another inteviewee reflected on the practical application of this theme in the Board's day-to-day operations: "Both the Board and the staff are looking for places where the Board can have an impact and make a policy advisory statement that will be used in setting a good precedent at Meta for human rights and for impacted peoples all over the world. And so, we're looking for [and] also considering representative cases. What are users most complaining about? Because this goes into the Board's considerations (…) So it's both what are people complaining about, and what's the potential impact on Facebook [and] Instagram enforcement practices, and then how those practices actually have a real-world impacting effect on people".

This theme also appers in media interviews. Board member Botero Marino: "One of the Board's fundamental tasks is to build a doctrine which is coherent and consistent with international human rights law (and, therefore, with democratic legal systems). It would help develop a doctrine that will serve judges and other operators when making decisions on moderation of digital online content" (Marino & Tuchtfeld, 2021).

And more subtly in this quote from Hunter-Torricke: "The board, though, exists to do something slightly different, which is to make decisions that Facebook has to comply with and come up with policy recommendations that Facebook has to consider and has to report back to us on" (Arenstein, 2021).

These two themes are supported by three practical strategies. The first is the Board's efforts to establish its professionalism and impartially. As an organization founded and funded by Meta, the Board met with skepticism by some of the company's critics and worries that it might be used as a so-called "fig leaf" (Ingram, 2020; Kabir, 2020b). The Board addressed these concerns by establishing a meticulous and transparent mode of operation, starting with the selection and training of its members. This included recruiting a prominent, professional and diverse group of people, which includes a Nobel Peace Prize laureate, the former Prime Minister of Denmark and former editors of notable newspapers such as The Guardian. Other members are distinguished law scholars, academics and experts who've held senior positions in public service. Then came a rigorous months-long training process, which included lessons in law and Meta's operation and mock tribunals. In addition, the members are supplemented with a large staff of legal aids, researchers, IT and media relations.

This strategy is also notable in the operation of Board's case selection team, as described by one interviewee: "We read as many as [of the cases] we possibly can, which is honestly not a million. We are able to sort of filter based on some basic characteristic of the appeals, like what part of the policy they're about, what language they're in. It's a bit of a I would say sampling, seeing what's out there on a number basis and then reading as many as you can in representatives slices of the cases (…) We have a diversity of expertise and background (…) we're trying to take good representative samples. And if there are cases or areas, we know that the board is not going to be interested in or know that the board is not going to be interested in, like right now because they're currently deliberating on the issue, then we might read less. But we still read it because we think it's important to know what's going on as best we can, throughout the whole queue of everything.

"The priorities of our team [and] just throughout the board, [are] professionalism, and careful analysis. This [is] the way that the administration's staff, with guidance from the board, operates. We want to be thorough about everything. That's kind of a universal value you'll find in the administration, staff (…) We're making big statements and big advice, so we want to be very, very careful and thorough when we examine any case, and I think that's a value that we have".

A completing principal of this strategy are the Board's efforts to maintain impartially throughout its process. "One of the overarching criteria is looking at the diversity of the cases, which is definitely considering geographic representation, because most of our [potential] cases actually come from the US", said an interviewee. "The Board is not wanting to be taking American cases all the time if it’s just took the luck of the draw. We are trying to distribute and make an impact in all of the different regions, as well as policies. The board wants to be able to comment on a bunch of different policies that may find impactful, important, and not just, for example, violence and incitement. And American cases the majority of cases we get now. But they wouldn't want to just take those kinds of cases all day long, there's a limit to how much usefulness they feel that would be. So, they try to distribute pretty carefully on the geographic diversity and the policy diversity (...) Those are the big things that we're constantly checking writ large (…) What's the mix of cases? Is it a good mix for the board having the best impact?”

The discussion and decision-making process, which is molti-staged and rigorous with numerous checks and decision points, is also geared towards professionalism and impartiality. First, the case is assigned to the Board’s computer system and all identifying data is striped. The Board than starts a preliminary stage which include questions and data requests from Meta and public comments. Inside researchers and outside institutes, the Board contracted gather relevant information, such as linguistic, political and cultural data. Only than does the Board begins discussions on the case. The first discussions are held in a committee of five Board members, at least one of them from the region related to the case. After a round of discussion one of the members volunteers to write a draft decision, with the help of research assistants. When done, all other committee members comment on the draft, and then meet again to reconcile differences. A final draft is written, which is then put up to the review of the full Board. All members can comment, discuss and suggest correction. Finally, the Board votes and publishes the decision. Final decision includes two main parts: decision on the specific case, and policy recommendations relevant to other similar cases. The entire process – from case selection to final decision – takes up to 90 days.

The second strategy focuses on the communication method with Meta. The Board is dependent on Meta to provide privileged information into its operation and decision-making process. Yet, over-dependence on Meta might hurt the Board’s credibility and impede its autonomy. It has dealt with this conundrum by treating Meta as a source of expert witnesses and knowledge with limited access to day-to-day operations. Meta provides the Board with relevant information, and when requested senior members will come and speak before the Board as part of its discussion. Other than that, there is no direct or constant communication. The company is not privileged to the Board’s discussions or internal communications, and members do not communicate with Meta’s staff. One interviewee stated: “We don’t see them, I never met them. At extraordinary events someone very senior will come from the company to answer questions (…) We did not meet Zuckerberg; we did not sit down to eat breakfast with him (…) They never read our emails. There is no email correspondence with them. We have a closed system, and they cannot be a part of our correspondence. There is no such option”.

In addition, the Board developed tools enabling direct communication with users, to reduce its dependence on Meta. “Once a case has been assigned to panel and it's being considered, it is possible to send a message over the Facebook user [who filed it] and ask them for more information”, said one Board perosnnel. “The user gets notifications as they go through the process so that they know that their content is being considered. The user goes to our website and authenticates that it's them through Facebook and Instagram and gives us permission to view their content and their user data, which is very important”.

The third strategy is direct acts of self-assertion, meant to preserve or enhance the Oversight Board’s autonomy. When confronted with instances where Meta tried to establish dominance or reduce the Board’s autonomy, its members and staff countered with forceful acts. A notable example: Following the events of the 6th of January 2021 in the US Capitol, and the ban placed by Meta of former US President Donald Trump's account, Meta referred the case to the Board for a final arbitration. Instead of issuing a definitive decision on the case, the Board proclaimed that it is up to Meta to develop at systematic policy and use it to decide on the case. "In applying a vague, standardless penalty and then referring this case to the Board to resolve, Facebook seeks to avoid its responsibilities. The Board declines Facebook's request and insists that Facebook apply and justify a defined penalty", wrote the Board in its decision (*Case Decision 2021-001-FB-FBR*, 2021).

By doing that, the Board defined the relationship with Meta according to its terms and prevented an incident where the company will have used the Board to evade responsibility. "[Meta] understood (…) that on the one hand it is fortunate that there are very serious and opinionated people there. It serves them. On the other hand, [Meta understood that] you can't mess with us", said a Board personnel.

One of the Board's co-chairs, former Prime Minister of Denmark Helle Thorning-Schmidt, asserted in a May 2021 media interview discussing the Trump decision: "We felt it was a bit lazy of Facebook to send over to us a penalties suggestion that didn't exist in their own rulebook, so to speak, or in their own community standards. And we are not here to lift responsibility of Facebook. We're here to be independent, to look at Facebook's own rules to ask whether they are following their own rules and issue advice standards they certainly want to follow. But we're not here to invent new rules for Facebook or take responsibility from Facebook to actually follow their own standards" (Allen, 2021).

Another example occurred during one of the Board's early training sessions held through a Zoom call. "I saw that there are a lot of non-video participants", said an interviewee. "I asked who these were, and people said they were from Facebook. We said we would not continue talking. They disconnected and did not return. There was a sensitivity that once we move on to a substantive thing, that they will not be present in the room".

The Oversight Board also did not shy away from confronting Meta on the breadth of its authority to select and review cases. In one of the Board's first cases, Meta reversed its decision to remove a post after the case was selected for review; claiming the Board should decline to hear the case as currently there no disagreement with the user. The Board rejected Meta's argument, stating that the need for disagreement applies at the moment a user exhausts the internal appeal process and adding that a decision to restore content does not render a case moot, explaining: "On top of making binding decisions on whether to restore pieces of content, the Board also offers users a full explanation for why their post was removed" (*Case Decision 2020-004-IG-UA*, 2021). Acquiescing to Meta in this case, the Board added, would enable Meta to effectively "exclude cases from review (…) integrate the Board inappropriately to Facebook's internal process and undermine the Board's independence" (*Case Decision 2020-004-IG-UA*, 2021). With this act, the Board established its authority to select cases and more importantly to interpret its jurisdiction and its charter (D. Wong & Floridi, 2022).

The Board also acted unilaterally to expand its oversight of Meta's handling of its recommendations, as Board member Julie Owono described in an October 2021 media interview: "At the beginning, in our bylaws, there was nothing about us tracking the impact of the work that we're doing and tracking the recommendation that we're making to the company. We have created an implementation working group, on which I'm currently sitting on, which is working on metrics to measure accurately, how Facebook is implementing our recommendations when they say they would implement them, and also to continue to push the company implement the recommendations that the company says it won't implement" (*Facebook Oversight Board Member on Transparency Report*, 2021).

The Board details its efforts to better monitor Meta in its first annual report, published June 2022. These efforts include creating and staffing a Case Implementation and Monitoring Team in July 2021, which "monitors and measures Meta's responses and actions, to understand the impact of our recommendations on Facebook and Instagram users" (Oversight Board, 2022b, p. 60). The Board later created an Implementation Working Group. This group, which includes Board Members and senior Meta staff, meets to discuss the recommendation process with Meta answering questions on the subject. As to the Implementation Committee of five Board members, referenced by Owono in the October 2021 interview, the report stated: "This represented a clear choice to place implementation on par with our organization's most critical functions" (Oversight Board, 2022b, p. 60).

In addition, the Board has developed a so-called "data-driven approach" to monitor how Meta implements its recommendations: "To measure Meta's progress on implementation, we looked at whether certain criteria for a given recommendation have been met (…) For other recommendations, Meta would need to provide data which isn't publicly available to demonstrate implementation (…) Going forward, we will measure Meta's implementation of our recommendations according to four categories, updating our assessments on a quarterly basis (…) Going forward, we will measure Meta's implementation of our recommendations according to four categories, updating our assessments on a quarterly basis (…) This new, data-driven approach means that our assessment of whether Meta has implemented a recommendation may at times differ from the company's reports" (Oversight Board, 2022b, pp. 60–61). With this approach, the Board further asserts it authority vis-à-vis Meta, expanding it role behind solely presenting recommendations which may or may not be followed, to one that includes monitoring and perhaps criticizing Meta's implementation of said recommendations or lack thereof. This role was not part of Meta's plan for the Board or part of its founding charter. The Board's decisions to take on this monitoring role helps it assert itself as an independent body with its own agenda and ideas, which might differ or contrast with those of Meta.

In these acts, the Board asserted itself as an independent critical entity, not afraid to confront Meta and act as a countering force. These acts of self-assertion further strengthen the Board autonomy, behind the formal scope of its bylaws.

The above three strategies partially correlate with the strategies known to be used by interdependent organizations to establish autonomy and mutual respect. These strategies were mapped by Wiedner & Mantere (2019) in a lognitudinal qualitative analysis of the English National Health service. The analysis was used to developed a model on how organizations spin off units while creating separate autonomous entities (Wiedner & Mantere, 2019). Though the Oversight Board is not a spin off from Meta, it is concerned with establishing its autonomy, and the strategies it uses corralate to the ones mapped by Wiedner & Mantere. Stratagies that are specifically relevant to this paper include developing and demonstrating competence, which incurs self-confidence and respect from the other entity; and cross-organisational communications, which is needed when a newly seperated and interdependent entity lack experties or resources and require access to relevant information. These two stratagies are interwined, and eventually lead to respect which enables autonomy. This, in turn, ingnites a ciclicle proces where the autonomy enhaches the entity's respect and vice-versa (Wiedner & Mantere, 2019).

The Oversight Board's focus on profesionalism and impartialy is an effort to demonstrate its competence, and hence worthiness, and to garner respect from Meta as well as external observares (i.e., media, regulators, politician, general public). “Developing and demonstrating competence generates appraisal respect” (Wiedner & Mantere, 2019, p. 672). In one example the focus of one organization’s employee “on evidence and rigor appeared to impress” members of the other organization (Wiedner & Mantere, 2019, p. 672). This is quiet similar to the approach which guides the Board in its operations, and might have a similar effect on Meta.

The second strategy correspondes with using communication as mean to facilitate respect, as constant communication as an important tool in the process. “Communicating across organizational boundaries facilitates developing and demonstrating competence”, and “Continued communication provides opportunities for parties to demonstrate their (growing) competence to one another and thus to make each other aware of it” (Wiedner & Mantere, 2019, pp. 670, 672). The Board’s relays on Meta for knowledge and expertise, and uses a conatant communications process which sources thoes expertise during its delibirations, thus enabling it to demonstrate its seriousness and throughness. This process helps foster mutual respect in both organizations. At the same time, the limited nature of the communication with Meta, and the Board’s efforts to establish direct and confidential communicatons with Meta’s users, signals the its autunomy from Meta.

The third strategy, direct acts of self assertion, does not corrolate to one of Wiedner & Mantere’s strategies, but can also be seen as a strategy ment to assert respect. While the first two strategies are more subtle or in-direct, the third builds on the Board’s self preception and is ment to overtly assert its autonomy and garner respect. The Board’s acts under this stratagy serves as a outright defience of Meta’s presived role for the Board, and a demonstration of its willingness to expande its authroty, even unilatarlly. This strategy signals Meta that the Board is not an entity under its control or influence. Meta’s response to the Board’s decision on Trump’s suspension, in which the company followed all of the its demands (Clegg, 2021), indicates that the stratagy achived its intended goals. This strategy might be of interest to further studies in the area. All three strategies can foster greater respect, which in turn enables claiming and granting autonomy (Wiedner & Mantere, 2019).

Missing from this discussion is how the Board's actions were precived by Meta. Some evidance – the response to the Trump ruling, or Meta's implamantion of the majority of the Board's policy recommandations (Clegg, 2021; Meta, n.d.) – indicate that the strategies are meeting with success. One Board personnel stated that a Meta request from the Board to provide "Policy Advisory Opinion" on two issues, as more evidance that Meta is accpeting its expanded role. Further study, and specifically interviews with relevant members of Meta's staff, is needed to make an informed determination.

Even so, we can detect a central theme to the Oversight Board's operation so far: the self-preceprion of its role as that of a policy creating entity, which is supported by practical stratgies comnon in interdependent organizations working to establish mutual respect and autonomy. These findings and their analysis provide sufficent evidance to accept the second hypothesis, and to state that it is by its actions Oversight Board's managed to established its respect and autonomy vis-à-vis Meta, leading the company to accept most of its recommendations.

The current findings indicate that this new type of regulatory intermediary is a reliable solution to content moderation issues on big social media plarforms. Members of the Oversight Board, though founded an funded by Meta and dependent on it, see it as an autonomus entity tasked with shaping content moderation policy at a high-level, employ stratagies the support this notion, and these strategies are meeting with success. The Board, as its members believe, managed to expand its roles behind those originally intend for it, and become an authrotive body shaping policy decision. Though further study and time is needed to assess the Board’s full impact, its role as a reliable regulatory solution can not be dismissed.

5. Conclusion

Content moderation on major social media plarforms is fraught with problems and pitfalls. Current solutions, which mostly relied on self-regulation by the platforms, have proved insufficient in the eyes of the politicians, regulators and the public. In effect, the platforms' content moderation efforts are considered determenental to a solution. In this paper, I hypothised that Meta's novel soultion to this issue – a coporation created reuglatory intermediary in the form of the Oversight Board – can form a reliable solution to content moderation issue.

To test this hypothesis I first analyzed the Board's formak autonomy. The analysis was based on the literature around political formal indepedence, and used a modifed version of indices developed by Gilardi and Hanretty & Koop (Gilardi, 2002, 2005; Hanretty & Koop, 2012; Koop & Hanretty, 2018) to score the formal autonomy of the Board on a scale of 0 to 1 based on indicators such as term length of Board members or budget control. The score was then converted into a literal description on a scale of low (level of autonomy) to high. This analysis indicated that the Board was created with a medium level of formal autonomy according to the modified Gilardi index, or a medium-high level of formal autonomy according to an expanded modified index incoporating indicators suggested by Hanretty & Koop. However, the indices failed to capture a few idiosyncrasies of the Oversight Board, such as the Board's power to amend its bylaws, to the status of Board members as part-time contract workers maintaining separate careers in other sectors. Viewed together, these unique institutional design features enhance the Board's formal autonomy by a considerable degree. This, however, is tempered by Meta ultimate control of the Board's funding and the lack of checks and balances in a private entity which allow it, at its discretion, to cut future funding.

To assess the Oversight Board's de facto autonomy, I used metrics based on its first 22 decisions finalized by the 1st of February 2022. First, I analyzed whether The Board concurred or overturned Meta's final decision. The Board overturned Meta's decision in 11 cases. However, of the remaining 11 cases Meta itself overturned its original decision after the Board selected the case for review. In all, the Board's process resulted in overturning 17 to Meta's decisions – indicating a high level of de facto autonomy. I then analyzed the Oversight Board's recommendations by assessing the burden their implementation will place on Meta. More than half of recommendations falls into one of the two lowest tiers of burden but represent a substantial cumulative effect and their implementation will result in considerable change to Meta's operations. In addition, the recommendations in the two highest tiers are still numerous and represent a considerable burden, with implementation that will require considerable resources to study or collect data and even some loss of autonomy on Meta's part. The conclusion is that the Oversight Board was designed with a high degree of formal autonomy and can operate with a similar high degree of de facto autonomy, enabling it to become a reliable critic of Meta.

Seeing as the Oversight Board can operate with autonomy, I turned to assess the possible impact its decisions have on Meta policy and operation. A review of Meta's response to the Board's recommendations, and the Board's assessment of their implementation, revealed that Meta implemented or made steps towards implementing 80.5% of them – a demonstrably high percentage indicating the Board is successful in affecting changes to Meta's operation. I than turned to examine what factors might influence Meta's willingness to accept the Board's recommendations in such high percentage and focus on to types of factors: internal factors, related to the Board's relationship with Meta, and external factors.

The examination of external factor was done through a quantitative method and focused on the possible influence of press coverage sentiment and stock price. Though an analysis of press coverage found that during 2021, the first full year of the Board's operation sentiment was significantly negative compared to 2019 and 2018, a further analysis did not provide any significant indication of a direct correlation between Meta's responses to the Board's recommendations and coverage sentiment or stock price.

To examine the internal factors, I used a qualitative approach based on interviews with Oversight Board personnel, media interviews with Board members and staff and documents review. The findings indicate that the Board personnel perceive it as an autonomous entity tasked with the influential role of shaping Meta's policies through in-depth recommendations and arbitration. They also operate to expand its roles behind the ones originally intended. This is apparent in two central self-image themes in the Board's personnel narrative: First, that the Board is an autonomous entity not beholden to Meta, and one that has grown behind the company's original intentions (the "Golem" that struck its maker, according to a Board personnel). Second, that the Board is not simply an arbiter on specific cases, but an entity which develops and promotes new policy solutions. At the same time, the members and staff are aware of their dependency on Meta and possible criticism that this dependency might create among outside observers. To counter this, the Board employs various strategies meant to guard its autonomy and garner respect. These include emphasis on professionalism in recruitment, training and case discussion; a communication strategy that is based on sourcing expert opinion from Meta while limiting its staff access to inside debate and correspondence, which enabled the Board access to valuable information while displaying competence and minimizing interference; and direct acts of self-assertion, most notably the Board's refusal to accept a role in the case of former US president

Donald Trump, which would have diminished its autonomy. The first two strategies are known in interdependent organization trying to assert autonomy, as mapped by Wiedner & Mantere (2019), reflecting the similar challenges faced by the Board and entities undergoing separation from a mother-organization.

The findings of this paper advance the understanding of the role of regulatory intermediaries by demonstrating that the model of the Oversight Board can be a reliable solution to mitigate content moderation problems in online user content platforms. Though dependent on Meta, the Board can operate with a considerable degree of autonomy, and its personnel is successfully using strategies to bolster its respect vis-à-vis Meta and enhance its autonomy. This, in turn, results in Meta adopting most of its recommendations and willing to changes the company's content moderation policy accordingly. The Board also expanded on its original role, and is considering itself as an entity tasked not only with spesific content moderation problems, but with developing a set of forward thinking ideas and solutions relevant to all social media platforms. The finding turns the Oversight Board model into a reliable template of enhanced self-regulation using regulatory intermediary that can be employed by other social media platforms and indeed other industries looking for viable self-regulation solutions. The findings also help in closing theoretical gaps in understanding the operation of regulatory intermediaries, demonstrating that they can be autonomus and impactfull enteties even when created by and somewhat dependent on corporation. In contrast, the paper illuminates new gaps in current theory and reseatch, emphesizing the need for a more nuanced understanding of regulatory intermediaries and the variuos types of intermediaries.

However, the findings are limited by paper's time frame. Though it establishes the Oversight Board can operate with autonomy and that Meta is willing to largely implement most of its recommendation, the paper leaves open the questions as to how the changes adopted by Meta to its policy and operation affect its content moderation efforts, and whether they are able to mitigate content moderation issues and the criticism surrounding them. This kind of analysis will require a longer timeframe, one which will allow policy and operational changes to have impact.

References

Abbott, K. W., Levi-Faur, D., & Snidal, D. (2017a). Theorizing Regulatory Intermediaries: The RIT Model. *The ANNALS of the American Academy of Political and Social Science*, *670*(1), 14–35. https://doi.org/10.1177/0002716216688272

Abbott, K. W., Levi-Faur, D., & Snidal, D. (2017b). Introducing Regulatory Intermediaries. *The ANNALS of the American Academy of Political and Social Science*, *670*(1), 6–13. https://doi.org/10.1177/0002716217695519

Allen, M. (2021, May 6). Watch: A conversation on the Facebook Oversight Board decision on Trump. *Axios*. https://www.axios.com/2021/05/03/axios-event-facebook-decision-trump

Amnesty International. (2022). *Myanmar: The social atrocity: Meta and the right to remedy for the Rohingya*. https://www.amnesty.org/en/documents/asa16/5933/2022/en/

Arenstein, S. (2021, May 5). A Chat with Dexter Hunter-Torricke, Lead Communicator, Facebook Oversight Board. *PRNEWS*. https://prnewsonline.com/facebook-oversight-board/

Arun, C. (2022). Facebook's Faces. *Harvard Law Review*. https://harvardlawreview.org/2022/03/facebooks-faces/

Atoklo, D. (2020, May 19). 'I have no intention of protecting Facebook' … Afia Asare-Kyei, Facebook's Oversight Board member pledges promotion of transparency and freedom of expression. *The Business & Financial Times*. https://thebftonline.com/2020/05/19/i-have-no-intention-of-protecting-facebook-afia-asare-kyei-facebooks-oversight-board-member-pledges-promotion-of-transparency-and-freedom-of-expression/

Ayres, I., & Braithwaite, J. (1992). *Responsive Regulation: Transcending the Deregulation Debate* (143766). Oxford University Press; eBook Academic Collection (EBSCOhost).

https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=143766&site=eds-live

Boyd, R., Ashokkumar, A., Seraj, S., & Pennebaker, J. (2022). *The Development and Psychometric Properties of LIWC-22*. University of Texas at Austin. https://doi.org/10.13140/RG.2.2.23890.43205

Braithwaite, J. (1993). Responsive business regulatory institutions. *Business Ethics and the Law*.

Brès, L., Mena, S., & Salles-Djelic, M.-L. (2019). Exploring the formal and informal roles of regulatory intermediaries in transnational multistakeholder regulation. *Regulation & Governance*, *13*(2), 127–140. https://doi.org/10.1111/rego.12249

Budzinski, O., & Mendelsohn, J. (2021). *Regulating Big Tech: From Competition Policy to Sector Regulation?* (SSRN Scholarly Paper ID 3938167). Social Science Research Network. https://doi.org/10.2139/ssrn.3938167

Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Campbell, A. J. (1998). Self-Regulation and the Media. *Federal Communications Law Journal*, *51*(3), 711–772.

Campbell, J. L. (2007). Why Would Corporations Behave in Socially Responsible Ways? An Institutional Theory of Corporate Social Responsibility. *Academy of Management Review*, *32*(3), 946–967. https://doi.org/10.5465/AMR.2007.25275684

Case Decision 2020-002-FB-UA, 2020-002-FB-UA (Oversight Board January 28, 2021). https://www.oversightboard.com/decision/FB-I2T6526K

Case Decision 2020-003-FB-UA, (Oversight Board January 28, 2021). https://www.oversightboard.com/decision/FB-QBJDASCV

Case Decision 2020-004-IG-UA, (Oversight Board January 28, 2021). https://www.oversightboard.com/decision/IG-7THR3SI1

Case Decision 2020-005-FB-UA, (Oversight Board January 28, 2021). https://www.oversightboard.com/decision/FB-2RDRCAVQ

Case Decision 2020-006-FB-FBR, (Oversight Board January 28, 2021). https://www.oversightboard.com/decision/FB-XWJQBU9A

Case Decision 2020-007-FB-FBR, 2020-007-FB-FBR (Oversight Board February 12, 2021). https://oversightboard.com/decision/FB-R9K87402/

Case Decision 2021-001-FB-FBR, 2021-001-FB-FBR (Oversight Board April 13, 2021). https://oversightboard.com/decision/FB-691QAMHJ/

Case Decision 2021-003-FB-UA, 2021-003-FB-UA (Oversight Board April 13, 2021). https://oversightboard.com/decision/FB-H6OZKDS3/

Case Decision 2021-006-IG-UA, 2021-006-IG-UA (Oversight Board July 8, 2021). https://oversightboard.com/decision/IG-I9DP23IB/

Case Decision 2021-009-FB-UA, 2021-009-FB-UA (Oversight Board September 14, 2021). https://oversightboard.com/decision/FB-P93JPX02/

Case Decision 2021-012-FB-UA, 2021-012-FB-UA (Oversight Board December 9, 2021). https://oversightboard.com/decision/FB-L1LANIA7/

Clegg, N. (2021). *In Response to Oversight Board, Trump Suspended for Two Years; Will Only Be Reinstated if Conditions Permit*. Meta. https://about.fb.com/news/2021/06/facebook-response-to-oversight-board-recommendations-trump/

Coffey, B. S., & Fryxell, G. E. (1991). Institutional ownership of stock and dimensions of corporate social performance: An empirical examination. *Journal of Business Ethics*, *10*(6), 437–444. https://doi.org/10.1007/BF00382826

Coglianese, C., & Mendelson, E. (2010). *Meta-Regulation and Self-Regulation* (SSRN Scholarly Paper No. 2002755). https://papers.ssrn.com/abstract=2002755

Communications Decency Act of 1996, no. Public Law No: 104-104 (1996). https://www.congress.gov/104/plaws/publ104/PLAW-104publ104.pdf

Constine, J. (2020, January 28). Toothless: Facebook proposes a weak Oversight Board. *TechCrunch*. https://techcrunch.com/2020/01/28/under-consideration/

Cukierman, A., Web, S. B., & Neyapti, B. (1992). Measuring the Independence of Central Banks and Its Effect on Policy Outcomes. *The World Bank Economic Review*, *6*(3), 353–398. https://doi.org/10.1093/wber/6.3.353

Cusumano, M. A., Gawer, A., & Yoffie, D. B. (2021). Can self-regulation save digital platforms? *Industrial and Corporate Change*, *30*(5), 1259–1285. https://doi.org/10.1093/icc/dtab052

Dal Bó, E. (2006). Regulatory Capture: A Review. *Oxford Review of Economic Policy*, *22*(2), 203–225. https://doi.org/10.1093/oxrep/grj013

De Silva, N. (2017). Intermediary Complexity in Regulatory Governance: The International Criminal Court's Use of NGOs in Regulating International Crimes. *The ANNALS of the American Academy of Political and Social Science*, *670*(1), 170–188. https://doi.org/10.1177/0002716217696085

Donadelli, F., & van der Heijden, J. (2022). The regulatory state in developing countries: Redistribution and regulatory failure in Brazil. *Regulation & Governance*, *n/a*(n/a). https://doi.org/10.1111/rego.12459

Douek, E. (2019). Facebook's Oversight Board: Move Fast with Stable Infrastructure and Humility. *North Carolina Journal of Law & Technology*, *21*(1), 1–78.

Dutt, R., Deb, A., & Ferrara, E. (2019). "Senator, We Sell Ads": Analysis of the 2016 Russian Facebook Ads Campaign. In L. Akoglu, E. Ferrara, M. Deivamani, R. Baeza-Yates, & P. Yogesh (Eds.), *Advances in Data Science* (pp. 151–168). Springer. https://doi.org/10.1007/978-981-13-3582-2_12

Dvoskin, B. (2022a). *Expert Governance of Online Speech* (SSRN Scholarly Paper No. 4175035). https://papers.ssrn.com/abstract=4175035

Dvoskin, B. (2022b). Expertise and participation in the facebook oversight board: From reason to will. *Telecommunications Policy*, 102463. https://doi.org/10.1016/j.telpol.2022.102463

Edwards, G., & Waverman, L. (2006). The Effects of Public Ownership and Regulatory Independence on Regulatory Outcomes. *Journal of Regulatory Economics*, *29*(1), 23–67. https://doi.org/10.1007/s11149-005-5125-x

Egan, E., & Beringer, A. (2018, April 18). Complying With New Privacy Laws and Offering New Privacy Protections to Everyone, No Matter Where You Live. *Meta*. https://about.fb.com/news/2018/04/new-privacy-protections/

Epstein, D., & Medzini, R. (2021). *The View from Above: Framing of Digital Privacy in Post Cambridge Analytica Congressional Hearings*. https://doi.org/10.2139/ssrn.3898331

Facebook oversight board member on transparency report. (2021, October 21). https://www.youtube.com/watch?v=wgilevJLZZM

Facebook, Social Media Privacy, and the Use and Abuse of Data | United States Senate Committee on the Judiciary. (2018, April 18). https://www.judiciary.senate.gov/meetings/facebook-social-media-privacy-and-the-use-and-abuse-of-data

Fernández, M. (2020). *Making the Digital Services Act Work for Consumers*. The European Consumer Organization.

Gamper-Rabindran, S., & Finger, S. R. (2013). Does industry self-regulation reduce pollution? Responsible Care in the chemical industry. *Journal of Regulatory Economics*, *43*(1), 1–30. https://doi.org/10.1007/s11149-012-9197-0

Ghosh, D. (2019, October 16). Facebook's Oversight Board Is Not Enough. *Harvard Business Review*. https://hbr.org/2019/10/facebooks-oversight-board-is-not-enough

Gilad, S. (2010). It runs in the family: Meta-regulation and its siblings. *Regulation & Governance*, *4*(4), 485–506. https://doi.org/10.1111/j.1748-5991.2010.01090.x

Gilardi, F. (2002). Policy credibility and delegation to independent regulatory agencies: A comparative empirical analysis. *Journal of European Public Policy*, *9*(6), 873–893. https://doi.org/10.1080/1350176022000046409

Gilardi, F. (2005). The Formal Independence of Regulators: A Comparison of 17 Countries and 7 Sectors. *Swiss Political Science Review*, *11*(4), 139–167. https://doi.org/10.1002/j.1662-6370.2005.tb00374.x

Gilardi, F., & Maggetti, M. (2011). The Independence of Regulatory Authorities. In D. Levi-Faur (Ed.), *Handbook on the Politics of Regulation* (pp. 201–214). Edward Elgar Publishing Limited. http://ebookcentral.proquest.com/lib/huji-ebooks/detail.action?docID=807373

Gillespie, T. (2017). Regulation of and by Platforms. In J. Burgess, T. Poell, & A. Marwick (Eds.), *SAGE Handbook of Social Media* (pp. 254–278). SAGE.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press.

Grabosky, P. (2013). Beyond Responsive Regulation: The expanding role of non-state actors in the regulatory process. *Regulation & Governance*, *7*(1), 114–123. https://doi.org/10.1111/j.1748-5991.2012.01147.x

Graves, S. B., & Waddock, S. A. (1994). Institutional Owners and Corporate Social Performance. *Academy of Management Journal*, *37*(4), 1034–1046. https://doi.org/10.5465/256611

Gunningham, N., & Rees, J. (1997). Industry Self-Regulation: An Institutional Perspective. *Law & Policy*, *19*(4), 363–414. https://doi.org/10.1111/1467-9930.t01-1-00033

Haggin, P. (2022, December 12). Elon Musk's Twitter Disbands Trust and Safety Council. *Wall Street Journal*. https://www.wsj.com/articles/elon-musks-twitter-disbands-trust-and-safety-council-11670898329

Hanretty, C., & Koop, C. (2012). Measuring the formal independence of regulatory agencies. *Journal of European Public Policy*, *19*(2), 198–216. https://doi.org/10.1080/13501763.2011.607357

Hartmann, I. A. (2022). Self-regulation in Online Content Platforms and the Protection of Personality Rights. In M. Albers & I. W. Sarlet (Eds.), *Personality and Data Protection Rights on the Internet: Brazilian and German Approaches* (pp. 267–287). Springer International Publishing. https://doi.org/10.1007/978-3-030-90331-2_11

Hawkins, K., & Hutter, B. M. (1993). The Response of Business to Social Regulation in England and Wales: An Enforcement Perspective*. *Law & Policy*, *15*(3), 199–217. https://doi.org/10.1111/j.1467-9930.1993.tb00103.x

Hensel, A. (2018, November 16). Facebook's 'independent oversight group' is destined to fail. *VentureBeat*. https://venturebeat.com/business/facebooks-independent-oversight-group-is-destined-to-fail/

Héritier, A., & Eckert, S. (2008). New Modes of Governance in the Shadow of Hierarchy: Self-regulation by Industry in Europe. *Journal of Public Policy*, *28*(1), 113–138. https://doi.org/10.1017/S0143814X08000809

Ingram, M. (2020, May 7). Facebook's new oversight board: Supreme Court or fig leaf? *Columbia Journalism Review*. https://www.cjr.org/the_media_today/facebooks-new-oversight-board-supreme-court-or-fig-leaf.php

Kabir, O. (2020a, May 7). Facebook Won't Fudge Content Oversight, Says Israeli Legal Expert. *CTECH by Calcalist*. https://www.calcalistech.com/ctech/articles/0,7340,L-3819032,00.html

Kabir, O. (2020b, May 12). Facebook Will Allow an Outside Body to Remove Content, but What Can't It Do? (In Hebrew). *Calcalist*. https://www.calcalist.co.il/internet/articles/0,7340,L-3821392,00.html

Kabir, O. (2021, April 12). Supreme Court: State Attorney Can Ask Social Networks to Remove Content (in Hebrew). *Calcalist*. https://www.calcalist.co.il/internet/articles/0,7340,L-3904254,00.html

King, A. A., & Lenox, M. J. (2000). Industry Self-Regulation without Sanctions: The Chemical Industry's Responsible Care Program. *The Academy of Management Journal*, *43*(4), 698–716.

Kjær, P., & Langer, R. (2005). Infused with news value: Management, managerial knowledge and the institutionalization of business news. *Scandinavian Journal of Management*, *21*(2), 209–233. https://doi.org/10.1016/j.scaman.2005.02.012

Klonick, K. (2020). *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression* (SSRN Scholarly Paper No. 3639234). https://papers.ssrn.com/abstract=3639234

Klonick, K. (2021, February 12). Inside the Making of Facebook's Supreme Court | The New Yorker. *The New Yorker*. https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court

Koop, C., & Hanretty, C. (2018). Political Independence, Accountability, and the Quality of Regulatory Decision-Making. *Comparative Political Studies*, *51*(1), 38–75. https://doi.org/10.1177/0010414017695329

Kourula, A., Paukku, M., Peterman, A., & Koria, M. (2019). Intermediary roles in regulatory programs: Toward a role-based framework. *Regulation & Governance*, *13*(2), 141–156. https://doi.org/10.1111/rego.12226

Kurtzleben, D. (2018, April 11). Did Fake News On Facebook Help Elect Trump? Here's What We Know. *NPR*. https://www.npr.org/2018/04/11/601323233/6-facts-we-know-about-fake-news-in-the-2016-election

Langvardt, K. (2017). Regulating Online Content Moderation. *Georgetown Law Journal*, *106*(5), 1353–1388.

Levi-Faur, D. (2011). Regulatory networks and regulatory agencification: Towards a Single European Regulatory Space. *Journal of European Public Policy*, *18*(6), 810–829. https://doi.org/10.1080/13501763.2011.593309

Levi-Faur, D., & Starobin, S. M. (2014). Transnational Politics and Policy: From Two-Way to Three-Way Interactions. *Jerusalem Papers in Regulation & Governance*, *62*, 1–38.

Lloyd, L. (2022, March 22). Facebook's Oversight Board's Work – And Other Free Speech Challenges | Bush Center. *George W. Bush Presidential Center*. http://www.bushcenter.org/publications/articles/2022/03/democracy-talks-mcconnell-facebook-oversight.html

Lytton, T. D. (2017). The Taming of the Stew: Regulatory Intermediaries in Food Safety Governance. *The ANNALS of the American Academy of Political and Social Science*, *670*(1), 78–92. https://doi.org/10.1177/0002716217690330

Mackintosh, E. (2021, October 25). Facebook knew it was being used to incite violence in Ethiopia. It did little to stop the spread, documents show | CNN Business. *CNN*. https://www.cnn.com/2021/10/25/business/ethiopia-violence-facebook-papers-cmd-intl/index.html

Maggetti, M. (2007). De facto independence after delegation: A fuzzy-set analysis. *Regulation & Governance*, *1*(4), 271–294. https://doi.org/10.1111/j.1748-5991.2007.00023.x

Maggetti, M. (2009). The role of independent regulatory agencies in policy-making: A comparative analysis. *Journal of European Public Policy*, *16*(3), 450–470. https://doi.org/10.1080/13501760802662854

Mahoney, L., & Roberts, R. W. (2007). Corporate social performance, financial performance and institutional ownership in Canadian firms. *Accounting Forum*, *31*(3), 233–253. https://doi.org/10.1016/j.accfor.2007.05.001

Majone, G. (2001). Two Logics of Delegation: Agency and Fiduciary Relations in EU Governance. *European Union Politics*, *2*(1), 103–122. https://doi.org/10.1177/1465116501002001005

Margolis, J. D., & Walsh, J. P. (2001). *People and Profits?: The Search for A Link Between A Company's Social and Financial Performance*. Psychology Press. https://doi.org/10.4324/9781410600622

Marino, C. B., & Tuchtfeld, E. (2021). Quasi-Judicial Oversight Mechanisms for Social Platforms – A Conversation with Catalina Botero Marino, Co-Chair of the Oversight Board –. *RuZ - Recht Und Zugang*, *2*(3), 254–262. https://doi.org/10.5771/2699-1284-2021-3-254

Medzini, R. (2021a). Credibility in enhanced self-regulation: The case of the European data protection regime. *Policy & Internet*, *13*(3), 366–384. https://doi.org/10.1002/poi3.251

Medzini, R. (2021b). Enhanced self-regulation: The case of Facebook's content governance. *New Media & Society*, 146144482198935. https://doi.org/10.1177/1461444821989352

Medzini, R., & Levi-Faur, D. (2022). *Self-Governance via Intermediaries: Credibility in Three Different Modes of Governance*.

Meet the Board | Oversight Board. (n.d.). Retrieved August 9, 2022, from https://oversightboard.com/meet-the-board/

Meta. (n.d.). *Oversight Board recommendations | Transparency Center*. Retrieved November 20, 2022, from https://transparency.fb.com/oversight/oversight-board-recommendations/

Meta. (2021, June 1). How Meta’s third-party fact-checking program works. *How Meta’s Third-Party Fact-Checking Program Works*. https://www.facebook.com/facebookmedia

Meta. (2022). *Meta Reports Third Quarter 2022 Results*. https://s21.q4cdn.com/399680738/files/doc_financials/2022/q3/Meta-09.30.2022-Exhibit-99.1-FINAL.pdf

Milmo, D. (2021, December 6). Rohingya sue Facebook for £150bn over Myanmar genocide. *The Guardian*. https://www.theguardian.com/technology/2021/dec/06/rohingya-sue-facebook-myanmar-genocide-us-uk-legal-action-social-media-violence

Ng, L. H. X., Cruickshank, I. J., & Carley, K. M. (2022). Cross-platform information spread during the January 6th capitol riots. *Social Network Analysis and Mining*, *12*(1), 133. https://doi.org/10.1007/s13278-022-00937-1

Nordlinger, E. (1987). Taking the state seriously. In M. Weiner & S. P. Huntington (Eds.), *Understanding political development* (pp. 353–390). Little, Brown.

Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate Social and Financial Performance: A Meta-Analysis. *Organization Studies*, *24*(3), 403–441. https://doi.org/10.1177/0170840603024003910

Oversight Board. (2020, October 22). *The Oversight Board is now accepting cases*. https://www.oversightboard.com/news/833880990682078-the-oversight-board-is-now-accepting-cases/

Oversight Board. (2021a, January 28). *Announcing the Oversight Board’s first case decisions*. https://www.oversightboard.com/news/165523235084273-announcing-the-oversight-board-s-first-case-decisions/

Oversight Board. (2021b, November). *Announcing the Board's next cases and changes to our Bylaws*. https://oversightboard.com/news/3138595203129126-announcing-the-board-s-next-cases-and-changes-to-our-bylaws/

Oversight Board. (2022a, February). *Oversight Board overturns Meta's original decision: Case 2021-015-FB-UA | Oversight Board*. https://oversightboard.com/news/327331385942204-oversight-board-overturns-meta-s-original-decision-case-2021-015-fb-ua/

Oversight Board. (2022b). *Oversight Board Annual Report 2021*. https://oversightboard.com/attachment/425761232707664/

Oversight Board. (2022c, July). *Securing ongoing funding for the Oversight Board*. https://www.oversightboard.com/news/1111826643064185-securing-ongoing-funding-for-the-oversight-board/

Oversight Board. (2022d). *Oversight Board Bylaws*.

Oversight Board. (2022e). *Oversight Board Q2 2022 transparency report*.

Oversight Board Charter. (2019). https://about.fb.com/wp-content/uploads/2019/09/oversight_board_charter.pdf

Pickup, E. L. (2021). The Oversight Board's Dormant Power to Review Facebook's Algorithms. *Yale Journal on Regulation Bulletin*, *39*, 1–22.

Porter, T., & Ronit, K. (2006). Self-Regulation as Policy Process: The Multiple and Criss-Crossing Stages of Private Rule-Making. *Policy Sciences*, *39*(1), 41–72. https://doi.org/10.1007/s11077-006-9008-5

Redrup, Y., & Tillett, A. (2019, March 28). Social media platforms can't self-regulate. *Australian Financial Review*. https://www.afr.com/technology/social-media-platforms-can-t-self-regulate-20190327-p517y5

Rees, J. (1988). *Reforming the Workplace: A Study of Self-Regulation in Occupational Safety*. University of Pennsylvania Press.

Robins-Early, N. (2021, November 8). How Facebook Is Stoking a Civil War in Ethiopia. *Vice*. https://www.vice.com/en/article/qjbpd7/how-facebook-is-stoking-a-civil-war-in-ethiopia

Rosenberg, M., Confessore, N., & Cadwalladr, C. (2018, March 17). How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

Schroepfer, M. (2018, April 4). An Update on Our Plans to Restrict Data Access on Facebook. *Meta*. https://about.fb.com/news/2018/04/restricting-data-access/

Schultz, M. (2021). Six Problems with Facebook's Oversight Board. Not enough contract law, too much human rights. In R. Broeml, J. Lüdemann, R. Podszun, & H. Schweitzer (Eds.), *Perspectives on Platform Regulation: Concepts and Models of Social Media Governance Across the Globe* (Vol. 1, pp. 145–164). Nomos Verlagsgesellschaft mbH & Co. KG. https://www.nomos-elibrary.de/10.5771/9783748929789-145/six-problems-with-facebook-s-oversight-board-not-enough-contract-law-too-much-human-rights?page=1

Self-regulation of social media platforms failing to curb disinformation, says new report. (2019, October 11). https://www.oii.ox.ac.uk/news-events/news/self-regulation-of-social-media-platforms-failing-to-curb-disinformation-says-new-report

Sinclair, D. (1997). Self-Regulation Versus Command and Control? Beyond False Dichotomies. *Law & Policy*, *19*(4), 529–559. https://doi.org/10.1111/1467-9930.00037

Smithey, S. I., & Ishiyama, J. (2000). Judicious choices: Designing courts in postcommunist politics. *Communist and Post-Communist Studies*, *33*(2), 163–182. https://doi.org/10.1016/S0967-067X(00)00002-7

Smyth, S. M. (2020). The Facebook Conundrum: Is it Time to Usher in a New Era of Regulation for Big Tech? *International Journal of Cyber Criminology*, *13*(2), 578–595. https://doi.org/10.5281/zenodo.3718955

Stigler, G. J. (1971). The Theory of Economic Regulation. *The Bell Journal of Economics and Management Science*, *2*(1), 3–21. https://doi.org/10.2307/3003160

Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Teoh, H. Y., & Shiu, G. Y. (1990). Attitudes towards corporate social responsibility and perceived importance of social responsibility information characteristics in a decision context. *Journal of Business Ethics*, *9*(1), 71–77. https://doi.org/10.1007/BF00382566

Waddock, S. A., & Graves, S. B. (1997). The Corporate Social Performance–Financial Performance Link. *Strategic Management Journal*, *18*(4), 303–319. https://doi.org/10.1002/(SICI)1097-0266(199704)18:4<303::AID-SMJ869>3.0.CO;2-G

Wiedner, R., & Mantere, S. (2019). Cutting the Cord: Mutual Respect, Organizational Autonomy, and Independence in Organizational Separation Processes. *Administrative Science Quarterly*, *64*(3), 659–693. https://doi.org/10.1177/0001839218779806

Wong, D., & Floridi, L. (2022). Meta's Oversight Board: A Review and Critical Assessment. *Minds and Machines*. https://doi.org/10.1007/s11023-022-09613-x

Wong, J. C. (2019, July 12). Facebook to be fined $5bn for Cambridge Analytica privacy violations – reports. *The Guardian*. https://www.theguardian.com/technology/2019/jul/12/facebook-fine-ftc-privacy-violations

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, *12*(4), 505–523. https://doi.org/10.1111/rego.12158

Appendices

Appendix I: Formal Autonomy Questionnaire & Answers

For each question, the selected answer is indicated in bold. Questions and answers suggested by Hanretty & Koop and used in the expanded index are underlined.

I: Status of the Oversight Board's Co-Chairs (weight 0.2)

1. Term of office
Over 8 years 1.00
6 to 8 years 0.80
5 Years 0.60
4 years 0.40
Fixed term under 4 years or at the discretion of the appointer 0.20
No fixed term 0.00

2. Who appoints the co-chairs of the Oversight Board? *
The members/co-chairs of the Oversight Board 1
The Board members/co-chairs and the trustees .75
The trustees .5
The trustees and meta .25
Meta 0

* While the first four co-chairs were selected by Meta, the bylaws formalized a process in which the outgoing co-chairs recommend replacements for the approval of the trustees

3. Are there appointment criteria for the position of Board co-chair?
Yes 1
No 0

4. Dismissal
Dismissal is impossible 1
Dismissal is possible, but only for reasons not related to policy 0.67
There are no specific provisions for dismissal 0.33
Dismissal is possible at the appointer’s discretion 0.00

5. Do the founding documents give provisions relating to the incompatibility of a Board co-chair?
Yes 1
No 0

6. May the co-chairs of the Board hold a position in Meta?
No 1

Yes, with the permission of Meta and/or the Oversight Board 0.5
Yes / no specific provisions 0

7. Is the appointment renewable?
No 1
Yes, once 0.5
Yes, more than once 0

8. Is independence a formal requirement for the appointment?
Yes 1
No 0

II: Status of the Oversight Board's Members (weight 0.2)

1. Term of office
Over 8 years 1
6 to 8 years 0.8
5 Years 0.6
4 years 0.4
Fixed term under 4 years or at the discretion of the appointer 0.2
No fixed term 0

2. Who appoints the members of the Oversight Board? *
The members/co-chairs of the Oversight Board 1
The Board members/co-chairs and the trustees .75
The trustees .5
The trustees and meta .25
Meta 0

* While the first slate of members was selected by the co-chairs and Meta, the bylaws formalized a selection process based the current members selecting new ones which are than formally appointed by the trustees

3. Are there appointment criteria for the position of Board member?
Yes 1
No 0

4. Dismissal
Dismissal is impossible 1
Dismissal is possible, but only for reasons not related to policy .67
There are no specific provisions for dismissal .33
Dismissal is possible at the appointer’s discretion 0

5. Do the founding documents give provisions relating to the incompatibility of a Board member?
Yes 1
No 0

6. May the members of the Board hold a position in Meta?
No 1
Yes, with the permission of Meta and/or the Oversight Board .5
Yes / no specific provisions 0

7. Is the appointment renewable?
No 1
Yes, once .5
Yes, more than once 0

8. Is independence a formal requirement for the appointment?
Yes 1
No 0

III: Relationship with Meta (weight 0.2)

1. Is the independence of the Oversight Board formally stated?
Yes 1
No 0

2. What are the formal obligations of the Oversight Board vis-à-vis Meta?
There are no formal obligations 1
Presentation of an annual report for information only .67
Presentation of an annual report that must be approved .33
The Oversight Board is fully accountable to Meta 0

3. Can Meta overturn the decisions of the Oversight Board? (This relates to decisions in specific cases, not policy recommendations)
No 1
Yes, with qualifications .5
Yes, unconditionally 0

IV: Financial and organizational autonomy (weight 0.2)

1. What is the source of the Oversight Board's budget?
Autonomically generated income 1

Both Meta and autonomically generated income .5
Only or primarily Meta 0

2. How is the budget controlled?
By the Oversight Board 1
By the Oversight Board's trust .75
By a third party .5
By both the Oversight Board/Oversight Board's trust and Meta .25
By Meta only 0

3. Which body decides on the Oversight Board’s internal organization?
The Oversight Board 1
The Oversight Board and the trust .75
The Trust .5
Both the Oversight Board/trust and Meta .25
Meta 0

4. Which body oversees the agency’s personnel policy (hiring and firing staff, deciding on its allocation and composition)?
The Oversight Board 1
The Oversight Board and the trust .75
The Trust .5
Both the Oversight Board/trust and Meta .25
Meta 0

V: Regulatory competencies (weight 0.2)

1. Who makes decisions regarding content moderation in Meta's platforms?
The Oversight Board only 1
The Oversight Board and another independent authority 0.67
The Oversight Board and Meta 0.33
The Oversight Board has only consultative competencies 0

Sources: Adapted from Gilardi (2002, 2005) and Hanretty & Koop (2012; Koop & Hanretty, 2018).

Appendix II: Poisson Regression Results

These are the results of the Poisson regression of Meta's sentiment coverage by year. Dataset and variables are identical to the ones used for the linear regression model.

```
Poisson regression                                      Number of obs =  4,030
                                                        Wald chi2(3)  =  26.69
                                                        Prob > chi2   = 0.0000
Log pseudolikelihood = -27804.933                       Pseudo R2     = 0.0042

------------------------------------------------------------------------------
             |               Robust
        Tone | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        year |
       2018  |   .0516383   .0257198     2.01   0.045     .0012284    .1020482
       2019  |   .1282452   .0273993     4.68   0.000     .0745436    .1819469
       2020  |   .0209651   .0293013     0.72   0.474    -.0364645    .0783947
             |
       _cons |   3.212491   .0208373   154.17   0.000     3.171651    3.253331
------------------------------------------------------------------------------
```

Hebrew Abstract

האוטונומיה של מתווכים רגולטוריים:

מטא והמועצה המפקחת כמקרה מבחן

עומר שילוני

תקציר בעברית

פעילותן של פלטפורמות מדיה חברתית גדולות, דוגמת פייסבוק, אינסטגרם, טוויטר, יוטיוב וטיקטוק, מייצרת אתגרים חדשים ומורכבים לחברה שבה הן פועלות. סוגיות כמו הפצת של תוכן שנאה, הסתה לאלימות, גזענות ותיאוריות קונספירציה, עומדות במוקד של דיון ציבורי, ובלבו טענות שהן גורם שותף לאירועים אלימים דוגמת הג׳נוסייד בבני הרוהינגה במיאנמר. סוגיות אחרות, כמו פרטיות, שימוש במידע אישי, פרסום ממוקד, ופגיעה במודל העסקי של כלי תקשורת מסורתיים עמדו גם הן במרכז דיון ציבורי, משפטי, פוליטי ורגולטורי. אתגרים אלו רלוונטיים ומהותיים היום כפי שהיו לפני כעשר שנים, ואף אחד מהפתרונות הרבים – פנימיים או חיצוניים – שהוצעו לא יצר פריצת דרך בהתמודדות אתם.

מטא, החברה האם של פייסבוק, אינסטגרם ו-ווטסאפ, מפעילה מגוון מאמצים להתמודדות עם אתגרי ניטור התוכן, שאת מרביתם ניתן להגדיר כרגולציה עצמית. הבולטים שבהם פנימיים – מערך של עובדי קבלן שאמונים על ניטור תכנים שסומנו על ידי משתמשים או מערכות אוטומטיות ופועלים בהתאם למערכת כללים מקיפה שמטא יצרה ומפתחת. אחרים בשיתוף פעולה וולונטרי עם גורמים חיצוניים, דוגמת ״מסלול מהיר״ שמאפשר לרשויות מדינה (למשל, פרקליטות המדינה של ישראל) להעביר למטא דיווחים על תוכן בעייתי לתפיסתן ולזכות למענה מואץ, או שיתוף פעולה עם גופי בדיקת עובדות שניתוחים שלהם מוצמדים לפוסטים רלוונטיים. פלטפורמות אחרות, ובפרט טוויטר ויוטיוב, מפעילות כלים דומים. העדר ההצלחה של מאמצי רגולציה עצמית אלו, והביקורת הגוברת על מדיניות התוכן של הפלטפורמות השונות, דחפו כמה ממשלות, האיחוד האירופי הבולטת שבהן, לקדם חוקים שיפקיעו מידי הפלטפורמות, ובהן מטא, את היכולת להסיר תכנים לפי שיקול דעתן הבלעדי.

ב-2018, הציעה מטא (שאז עוד פעלה תחת השם פייסבוק) פתרון חדש כמענה לאתגרי ניטור התוכן שלה, והביקורת הציבורית המרובה שקשורה אליהם. הפתרון מנסה לצעוד בתלם שבין הקצוות הללו, רגולציה עצמית טהורה מחד ורגולציה ממשלתית מאידך, ולהעביר חלק מסמכויות בקרת התוכן לישות חדשה בשם ״המועצה

המפקחת" (the Oversight Board) גוף חיצוני עצמאי, אך בעל קשרים למטא, שישמש כמעין "בית משפט עליון" לתוכן, יוכל לקבל החלטות מחייבות בערעורים של משתמשים על החלטות מטא להסיר או להשאיר תוכן ולפרסם המלצות מדיניות לא מחייבות רחבות שנועדו להדריך ולעצב את מדיניות ניטור התוכן של מטא.

המועצה המפקחת החלה לפעול באוקטובר 2020. עד עתה, המחקר האקדמי אודותיה נכתב רובו ככולו על ידי משפטנים ועסק בסוגיות משפטיות של פעילותה, דוגמת הסתמכותה על מסגרת של כללי המשפט הבינלאומי לזכויות אדם למסגור הדיונים. אף שמדובר בראש ובראשונה בפתרון רגולטורי לבעיית מדיניות, לא נעשה מחקר מעמיק שבוחן את פעילות המועצה מנקודת המבט של תחום המדיניות הציבורית. מחקר זה ממלא פער זה, ובוחן את פעילות המועצה מנקודת המבט התיאורטית והמעשית של תפקודה כמתווך רגולטורי.

מתווכים רגולטוריים הם גופים או מומחים שמספקים סיוע לרגולטורים בהוצאה לפועל של יעדיהם ביחס לגופים המבוקרים. הם יכולים להיות גורמים פנימיים בארגון – למשל, מבקרים פנימיים, קציני פיקוח ויועצים משפטיים – אך פעמים רבות מדובר בגופים חיצוניים אוטונומיים. מתווכים רגולטוריים יכולים להגיע מהשוק הפרטי, כמו חברות דירוג וסרטיפיקציה שפועלות למטרות רווח, משרדי ראיית חשבון, או סוכנויות דירוג אשראי; או להית ארגוני חברה אזרחית, ארגונים בין-ממשלתיים ואפילו מדינות שפועלות לקדם ציות של מדינות אחרות בהתאם למנדט ממועצת הביטחון של האו"ם. פעילותם של מתווכים רגולטוריים לא מוגבלת רק לפעילות של סוכנויות רגולציה מדינתיות, וכוללת סרטיפיקציה ותיוג של מוצרים, דיווח על ציות, דירוג מוצרים וחברות, בחינת ביצועים, ניטור ביצועים, דיווח על התנהגות לא נאותה וביקורת חיצונית על ארגונים. דוגמאות בולטות הן מפקחים פרטיים שמסייעים למנהל המזון והתרופות האמריקאי (FDA) לפקח על מזון מיובא, גופים כשרות בחו"ל שמעניקים הכשר למוצרי מזון, גופים שמנהלים רישום בעלים ומפיקים של כימיקלים מסוכנים, גופים שמדרגים את צריכת האנרגיה של מוצרי אלקטרוניקה, והשימוש שעושה האיחוד האירופי ברשתות בין-ממשלתיות של סוכנויות לאומיות כדי להטמיע בצורה עקבית את כללי האיחוד ולספק משוב. בחינה של פעילותם של מתווכים רגולטוריים ומערכות היחסים שלהם עם גורמי ממשל ורגולטורים (rule-makers) ועם מושאי פיקוח (rule-takers) מספקת הבנה חשובה על האופן שבו הם יכולים לשרת אינטרסים ציבוריים, ושל נקודות החוזקה והחולשה של פעילותם.

המקרה של המועצה המפקחת מספק הזדמנות לבצע בחינה זו דרך פריזמה ייחודית. המועצה היא מתווכת רגולטורית יוצאת דופן שמנהלת מערכת יחסים דואלית עם מטא. מהצד האחד, מטא היא הגורם שהקים את המועצה, מתקצב אותה ועיצב וקבע את גזרת הפעילות שלה (דרך המסמך המכונן של המועצה – Oversight Board Charter). מעבר לשלב ההקמה אמנם אין למטא תפקיד במינוי חברי המועצה המפקחת –

אלו שתפקידם לדון בתיקים, להכריע ולגבש המלצות מדינית – אך היא ממנה את חבר הנאמנים של המועצה (Trust), ששולטים בהקצאת וביישום התקציב שלה; דבר שמקנה לה שליטה, גם אם עקיפה ומרוחקת, על המועצה. ככזאת, היא ממלאת את תפקיד הדרג הפוליטי שמחליט על הקמת סוכנות רגולטורית ותווה את פעילותה. מנגד, מטא גם נתונה לביקורת המועצה: היא מחויבת להעביר לה מידע רלוונטי בתיקים שבהן דנה, למלא אחרי החלטותיה להסיר, להחזיר או להותיר תוכן בתיקים ספציפיים להגיב בתוך תקופה קצובה בזמן להמלצות המדיניות שלה. בפועל, אף שאלו אינן מחייבות, מטא גם מיישמת את הרוב הגדול של המלצות המדיניות. בהקשר זה, היא מושא פיקוח של המועצה בתפקידה כמתווכת רגולטורית. בחינת הפעילות של המועצה המפקחת דרך פרספקטיבה זו תאפשר למלא פערים תיאורטיים בהבנת פעילותם של מתווכים רגולטוריים, ובפרט בהיבטים כמו מידת האוטונומיה שלהם והיכולת שלהם להשפיע על ולעצב את התנהגות הגוף המבוקר.

את הפתרון של המועצה המפקחת אפשר גם לשייך לסט הפתרונות של רגולציה עצמית, הגם שבמקרה הנוכחי מדובר בפתרון שבו הגבולות בין הסביבה הפנימית לחיצונית מטושטשים, כשהמועצה היא מעין פתרון ביניים בין רגולציה עצמית לחיצונית. היעילות של רגולציה עצמית, בפרט של פלטפורמות תוכן גדולות ותאגידים בינלאומיים, שנויה במחלוקת והיא יכולה לשמש לא פעם ככסות שנועדה להרחיק או למתן רגולציה חיצונית, ולאפשר לתאגיד לצמצם שקיפות ולשמר שליטה בסוגיות שחשובות בעבורו. בהקשר זה, המועצה המפקחת משמשת מקרה בוחן לסוג חדש של רגולציה עצמית ובחינה של פעילותה יכולה לסייע להשיב על השאלה האם רגולציה עצמית מסוג זה יכולה לפעול, ובאילו תנאים.

ממצאי המחקר יספקו תובנות חשובות לא רק בהקשר של רגולציית תוכן או רגולציה של פלטפורמות תוכן גולשים, והם יכולים להיות רלוונטיים גם לתעשיות אחרות. תעשיית הכימיקלים, לדוגמה, עושה שימוש במודל רגולציה עצמית ברמת התעשייה, שמופעל על ידי הארגון המייצג Chemical Manufacturers Association. ואולם, מחקרים העלו שמודל זה אינו אפקטיבי, מאפשר התנהגות אופורטוניסטית, ושמפעלים שמסתמכים עליו מזהמים יותר מאשר מפעלים שלא אימצו את המודל. אישוש ההיתכנות של מודל המועצה המחוקקת יכול לספק תמריץ לתעשיות דוגמת תעשיית הכימיקלים לאמץ מודל דומה, כזה שיימנע מהכשלים של מודל שמופעל על ידי התעשייה עצמה.

מחקר זה מתמקד בבחינת מערכת היחסים הרשמית והמעשית בין מטא למועצה המפקחת, בניסיון להבין האם גוף שכזה יכול להיות פתרון בר-קיימא לסוגיית ניטור תוכן, ובהקשר רחב יותר גם לבעיות רגולציה

אחרות שבהן האינטרס הציבורי מאפשר או לעיתים מחייב מידה של רגולציה עצמית. זאת, תוך בחינה של שתי שאלות מחקר:

1. מה התנאים שבהם מתווך רגולטורי שנוצר וממומן על ידי ישות תאגידית יכול לפעול תוך אוטונומיה רחבה מספיק כדי להפוך לגורם מבקר אמין ומשפיע על אותה ישות?

על מנת לענות על שאלה זו בחנתי את המבנה המוסדי (institutional design) של המועצה המפקחת בהתייחס לממדים אמפיריים שונים של אוטונומיה, וכן את הפעולה האוטונומית הלכה למעשה של המועצה מול מטא. המסקנות העולות מן הניתוח הן שהמועצה פועלת ברמה גבוהה של אוטונומיה. לאור זאת, בחנתי את מידת ההיענות של מטא להמלצות המדיניות שמציבה לה המועצה, ובחינה העלתה שהחברה מגלה הענות גבוהה מאוד. לכן המשכתי ובדקתי את שאלת המחקר השנייה:

2. מהם הגורמים, חיצוניים או פנימיים, שמשפיעים על מידת ההיענות של מטא להחלטות והמלצות של מתווך רגולטורי פרי יצירתה?

על מנת לענות על שאלה זו ביצעתי ניתוח כמותני של הגורמים החיצוניים הבאים: סנטימנט הסיקור של מטא בעיתונות ומחיר המניה של מטא, וביצעתי רגרסיה לוגיסטית מולטילינארית כדי לבדוק האם הם מנבאים את תגובת מטא להמלצות המועצה. בחינה זו לא העלתה ממצאים מובהקים. כן ביצעתי ניתוח איכותני של הפעולות שבהן נוקטת המועצה כדי לרכוש את הכבוד של מטא ולהרחיב את האוטונומיה שלה. ניתוח זה העלה שהמועצה משתמשת בפעולות שנפוצות בארגונים בעלי תלות-הדדית לביסוס כבוד ואוטונומיה, ולכן ניתן להסביר את החלטת מטא לקבל את מרבית המלצותיה בגורמים פנימיים.

בהרחבה, כדי להשיב על שאלות המחקר התמקדתי בבחינת המידה והתנאים שבהם:

1. מטא יצרה את המועצה המפקחת כיישות בעלת מידה גדולה של אוטונומיה.
2. המועצה המפקחת מצליחה לפעול במידה גבוהה של אוטונומיה דה פקטו.
3. ההיענות של מטא להמלצות המועצה המפקחת נעוצה בגורמים חיצוניים שלא קשורים ישירות למערכת היחסים בין שני הגופים.
4. ההיענות של מטא להמלצות המועצה המפקחת נעוצה בגורמים פנימיים במערכת היחסים בין שני הגופים, ובאסטרטגיות שנוקטת המועצה על מנת לבסס את האוטונומיה שלה.

חלקו הראשון של המחקר מתמקד בבחינת השערות 1 ו-2, ובוחן את האוטונומיה של המועצה המפקחת דרך המערכת היחסים הדואלית שלה עם מטא. ראשית, נבחנת מערכת היחסים עם מטא בתפקיד הדרג הפוליטי שמקים סוכנות רגולטורית. בחינה זו מתבססת על הספרות המחקרית שבודקת אוטונומיה פוליטית של

סוכנויות רגולטוריות, ועושה שימוש באינדקסים שפותחו לצורך מחקר כמותני השוואתי בין סוכנויות שונות, ובפרט האינדקס של פבריציו ג'ילרדי שפותח על מנת להעריך את האוטונומיה הפורמלית של סוכנויות רגולטוריות. האינדקס מדרג את רמת האוטונומיה של סוכנות בסולם של 0 עד 1, כאשר 0 מעיד על העדר אוטונומיה ו-1 על אוטונומיה מלאה. אינדקס זה עבר התאמה למקרה הייחודי של המועצה המפקחת ומטא, ולכן לא ניתן להשתמש בו בהיבט השוואתי, במקום זאת הומר הציון לסולם הערכה מילולי להלן: 0-0.2, אוטונומיה נמוכה; 0.21-0.4, אוטונומיה נמוכה-בינונית; 0.41-0.6, אוטונומיה בינונית; 0.61-8 אוטונומיה בינונית-גבוהה; 0.81-1, אוטונומיה גבוהה. הדירוג מבוסס על שאלון בן חמישה חלקים שבוחן היבטים שונים של אוטונומיה. השאלון מולא על ידי נציג.ה מהצוות האדמיניסטרטיבי של המועצה המפוקחת, בתוספת מידע משלים ממקורות פומביים. האינדקס המתואם של ג'ילרדי העניק למועצה ציון של 0.6, שמעיד על אוטונומיה בינונית. שאלון מורחב יותר, שמשלב הצעות של הנרטי וקופ, העניק למועצה ציון של 0.64, שמעיד על אוטונומיה בינונית-גבוהה.

עם זאת, האינדקס מספק רק תמונה חלקית לאור מאפיינים ייחודיים של המועצה המפקחת, שמגבירים את האוטונומיה שלה באופן שלא מבוטא בשאלון. אלו כוללים בחירת חברי מועצה שמתחזקים קריירות עצמאית נפרדות ובולטות בתחומים כמו אקדמיה, תקשורת, מגזר שלישי ופוליטיקה. חברי המועצה מועסקים בה רק במשרה חלקית, ובמקביל לקריירות המקצועיות שלהם. הם לא תלויים במועצה, ולכן גם בעקיפין במטא כמי שמתקצבת אותה, לפרנסתם או למוניטין המקצועי שלהם (החברים כוללים חוקרים בולטים באקדמיה, עורכי עיתונים ארציים גדולים, וגם ראשת ממשלה לשעבר וכלת פרס נובל), מה שמקנה להם אוטונומיה וחופש פעולה רחבים יותר. בנוסף, חברי המועצה יכולים לשנות לפי שיקול דעתם את הכללים שמסדירים את פעילותה (bylaws), ולהרחיב את היקף פעילותה, ואכן עשו זאת בכמה הזדמנויות. בשילוב מאפיינים ייחודיים אלו, שמעצימים את האוטונומיה של המועצה, ניתן לקבוע שהיא פועלת במידה רחבה של אוטונומיה פוליטית.

בחינת הצד השני של המטבע, מערכת היחסים על המועצה עם מטא בכובעה כמושא פיקוח, נעשתה דרך נקודת המבט של אוטונומיה דה פקטו, קרי האם החלטות המלצות המועצה משרתות את האינטרסים של מטא. הבחינה בוצע באמצעות שני מדדים. מדד ההחלטות בחן באילו מקרים המועצה קיבלה או דחתה את החלטות מטא. מדד ההמלצות מעריך את הנטל שהמלצות המדיניות השונות מטילות על מטא. הבחינה מקיפה את 22

ההחלטות הראשונות של המועצה,[8] שהתפרסמו בין ה-28 בינואר 2020 ל-1 בפברואר 2022, ומייצגת לכן את פעילותה בשנתה הראשונה.

במדד ההחלטות, המועצה הפכה את ההחלטה הסופית של מטא ב-11 מתוך 22 מקרים בלבד. ואולם, מתוך 11 המקרים שבהם הסכימה המועצה עם מטא, ב-6 מהם מטא עצמה שינתה את ההחלטה המקורית שלה לאחר שהתיק נבחר לדיון על ידי המועצה, וכתוצאה מהליך בדיקה פנימי נוסף שערכה מטא. בפועל, ההליך של המועצה הביא להפיכת 17 מ-22 ההחלטות המקוריות של מטא. העובדה שבכרבע מהמקרים עצם הבחירה בתיק לדיון הובילה את מטא לשנות את עמדתה המקורית, ושמבין המקרים הנותרים המועצה דחתה את החלטה של מטא במרבית המקרים מצביעה על רמה גבוהה של אוטונומיה דה פקטו.

לצורך גיבוש מדד ההמלצות זוהו תחילה 97 המלצות מדיניות ייחודיות שנכללו ב-22 ההחלטות הראשונות של המועצה. המלצות אלו מופו לתוך אחת מ-10 קטגוריות בהתאם לאופיין (שינוי מדיניות, שינוי תפעולי וכו׳). קטגוריות אלו חולקו לאחת מארבע קבוצות היררכיות, בהתאם לנטל שיישומן יטיל על מטא בהיבטים כמו הקצאת משאבים או פגיעה באוטונומיה, כאשר קבוצה במספר גדול יותר מייצגת נטל רב יותר. זאת, על סמך ההסקה של המלצות שמחייבות את מטא לנטל גדול יותר מייצגות מידה גדולה יותר של אוטונומיה דה פקטו.

מבין ההמלצות שמופו 10.3% סווגו לקבוצה 1, שמייצגת את הנטל הנמוך ביותר, 48.5% לקבוצה 2, 23.7% לקבוצה 3 ו-17.5% לקבוצה 4. יוצא שרוב ההמלצות משתייכות לרף הנטל הנמוך יותר, והטמעתן פשוטה יחסית או לא מחייבת משאבים ניכרים, מה שיכול להעיד על רמה נמוכה של אוטונומיה דה פקטו. ואולם, להמלצות אלו יש השפעה מצטברת משמעותית ויישום כולן יניע שינוי משמעותי בפעילות השוטפת של החברה בהיבטים כמו נגישות המשתמשים להליכי ערעור והליכי אכיפת מדיניות התוכן של מטא. בנוסף, ההמלצות בשתי הקבוצות ברף הגבוה יותר עדיין מייצגות נתח משמעותי – 41.2% - מכלל המלצות המועצה ויישומן כרוך בנטל ניכר. בפרט ניתן לזהות מספר משמעותי של המלצות שדורשות ממטא לערוך ולפרסם מחקרי רוחב או לגבש מדיניות בתחומים שונים, המלצות שדורשות ממנה לבצע שינויי מדיניות שהוגדרו על ידי המועצה או לאפשר עריכת ביקורת מצד גופים חיצוניים ועצמאיים. במשותף, יישום המלצות אלו כרוך בנטל משמעותי. ניתוח זה מצביע על כך שהמועצה המפקחת פועלת כגורם מבקר אמין של מטא, שמבקש ממנה להטמיע שינויים רבים ומשמעותיים בפעילותה, ולכן מצביע על פעילות ברמה גבוהה של אוטונומיה דה פקטו.

[8] בפועל דנה המועצה ב-23 תיקים בתקופה זו, אך בתיק הראשון הפוסט שפרסם המשתמש נמחק על ידי אחרי בחירת התיק לדיון אך קודם להשלמת הדיונים, ולכן לא גיבשה המועצה החלטה בתיק.

חלקו השני של המחקר מתמקד בבחינת השערות 3 ו-4. ראשית, נבדק באיזה היקף מקבלת מטא את המלצות המועצה המפקחת. לפי נתוני המועצה המפקחת מאוקטובר 2022, מתוך 118 המלצות שמטא הגיבה להן, מטא קיבלה ויישמה בצורה מלאה או חלקית 80.5% מההמלצות. נתונים אלו מצביעים על כך שהמועצה מצליחה להשפיע על מדיניות מטא. אך מה הגורמים לכך?

הבסיס התיאורטי לבחינת השערה 3 הוא הספרות המחקרית שעוסקת בגורמים החיצוניים שמשפיעים על תאגידים לנקוט בהתנהגות אחראית חברתית. אחד הגורמים שספרות זו מצביעה עליה הוא התקשורת כמוסד חיצוני עצמאי שמבקר את ומפקח על תאגידים, ושדרך פעילותו יכולים להיווצר לחצים על תאגידים לנקוט בהתנהגות אחראית חברתית. גורם אחר שזוהה הוא ביצועים פיננסיים, ובפרט מחיר המניה של תאגיד ציבורי: ירידה במחיר המניה יכולה להוות תמריץ לארגונים לנקוט בהתנהגות אחראית חברתית על מנת לרצות משקיעים ולהפיג לחצים חיצוניים אחרים (כמו ביקורת רגולטורית או מצד הדרג הפוליטי). שני גורמים אלו הם גם מדדים מתווכים לגורמים אחרים. למשל, ביקורת רגולטורית מוגברת יכולה להוביל לסיקור תקשורתי ביקורתי יותר או לפגוע במחיר המניה של התאגיד.

בחינה השפעה של גורמים אלו נעשתה בשיטה כמותית. לבחינת ההשפעה האפשרית של סיקור תקשורתי הורדו ממאגרי מידע אקדמאיים כל הכתבות שעסקו במטא או בפלטפורמות התוכן שלה, ושפורסמו בניו יורק טיימס, בוול סטריט ג׳ורנל, בפייננשל טיימס ובאסוסייטד פרס בין 2018 ל-2021. הכתבות עובדו באמצעות תוכנת הניתוח הסמנטי LIWC-22, שמזהה בטקסט מלים בעלות סנטימנט חיובי או שלילי, ומדרגת את הטון הכולל בסולם של 1 עד 100 (1 שלילי לחלוטין, 100 חיובי לחלוטין). ניתוח זה העלה שב-2021 הטון הממוצע של הכתבות שעסקו במטא היה שלילי יותר מבין כל אחת מהשנים האחרות (אם כי, ככלל הסיקור התקשורתי את מטא מאופיין בסנטימנט שלילי). רגרסיה לינארית עם סנטימנט הסיקור כמשתנה תלוי השנה כמשתנה דמי עם 2021 כקבוצת היחס זיהה הבדל מובהק בין 2021 ל-2018 ול-2019. ממצא זה מעלה אפשרות שהסיקור התקשורתי השפיע על הנכונות של מטא לקבל את המלצות, ואולם דרוש ניתוח סטטיסטי נוסף על מנת לבחון קשר ישיר יותר בין הגורמים. לפני ניתוח זה נבחנה השפעת מחיר המניה. לצורך כך נוצר קובץ נתונים שמבוסס על מחיר המניה של מטא בסוף יום המסחר, בכל יום מסחר בין 2018 ל-2021. נתונים אלו העלו שב-2021 מחיר המניה היה גבוה יותר מכל שנה אחרת שנבדקה. ממצאים אלו מצביעים על כך שמחיר המניה לא היה גורם משפיע, בלי תלות בהתאם קיים הבדל מובהק בין השנים.

על מנת לבחון קשר ישר בוצעה רגרסיה מולטינומינאלית שבה המשתנה התלוי הוא תגובות מטא להמלצות המועצה (קבלה מלאה, קבלה חלקית, בדיקת היתכנות, המלצה שכבר מיושמת בפועל ודחייה) בין

פברואר 2021 לאוגוסט 2022, עם דחייה כקטגוריית היחס וכשהודעת המלצה שכבר מיושמת הושמטה מכיוון שאין בה רכיב של החלטה מצד מטא. המשתנה המסביר הראשון היה ממוצע סנטימנט הסיקור התקשורתי של מטא ב-30 הימים שקדמו להחלטה (בשנה הראשונה לפעילות המועצה הוקצבו למטא 30 יום להגיב לכל המלצה). המשתנה המסביר השני היה מחיר המניה הממוצע בתקופת 30 הימים. המשתנים המפקחים היו גם גיל המועצה בשבועות ביום פרסום ההחלטה של מטא ומספר הכתבות שעסקו במטא בתקופת 30 הימים. בניתוח זה, לשום מקדם לא היתה השפעה מובהקת. ניתוח אחר, רגרסיה לוגיסטית שבה כל הקטגוריות במשתנה התלוי שבהן מטא קיבלה בצורה מלאה, חלקית או הסכימה עקרונית להמלצות המועצה אוחדו לקטגוריה אחת ודחייה נותרה קטגוריית היחס, כאשר המשתנים הבלתי-תלויים לא השתנו, לא סיפקה תוצאות שונות מבחינת המובהקות של המקדמים.

אין לפיכך נתונים שמאפשרים לתמוך בהשערה 3, אם כי יש לסייג זאת לאור הכמות הקטנה של התצפיות, ובעיקר הכמות הקטנה של התצפיות בקטגוריית היחס. ייתכן שניתוח דומה שיבוסס על מספר תצפיות גדול יותר יעלה ממצאים מובהקים.

השערה 4 נבחנה בשיטה איכותנית, על בסיס ראיונות עומק עם אנשי המועצה המפקחת, ניתוח ראיונות שנתנו חברי המועצה לכלי תקשורת וניתוח מסמכים פומביים של המועצה. זאת, במטרה למפות אסטרטגיות שבהן עושה המועצה שימוש על מנת לבסס את הכבוד ואת האוטונומיה שלה ביחס למטא. ניתוח זה זיהה שתי תפיסות זהות עצמית של המועצה. הראשונה, ראיית המועצה כישות אוטונומית שלא פועלת מתוך מחויבות לצרכים של מטא. השנייה, תפיסת המועצה כגוף שאמור לשמש כמאיץ לפיתוח ויישום של רעיונות ופתרונות מתקדמים לסוגיות שקשורות למדיניות בקרת תוכן. תפיסות אלו נתמכות על ידי שלוש אסטרטגיות מעשיות: הראשונה, ביסוס המועצה כגוף מקצועי ונטול פניות; השנייה, פיתוח מערכת יחסים עם מטא שמבוססת על אסטרטגיית תקשורת שבה מטא משמשת כמקור מידע וידע מקצועי, אך לא כגורם שמעורב בדיוני המועצה הוא מתווך בינה לבין משתמשים; ושלישית, מעשים ישירים של התרסה וביסוס עצמי (self-assertion). אסטרטגיות אלו מקבילות חלקית לאסטרטגיות מוכרת מתחום המחקר הארגוני כאסטרטגיות שמשמשות ארגונים בעלי תלות-הדדית לביסוס יחסי כבוד ואוטונומיה הדדיים. ניתוח זה מצביע על כך שזוהי פעילות המועצה והמפקחת ומערכת היחסים שפיתחה עם מטא שמשמשת כגורם שמוביל את החברה לקבל את מרבית המלצותיה.

המחקר מעלה שהמועצה המפקחת פועלת ברמה גבוהה של אוטונומיה, ושבאמצעות אסטרטגיות שנקטה בהן הצליחה לבסס כבוד ולהעצים את האוטונומיה שלה ביחסיה עם מטא, באופן שמגדיל את נכונות החברה לקבל את המלצותיה. ממצאים אל מצביעים על כך שמתווך רגולטורי דוגמת המועצה יכול לשמש

כפתרון אמין לסוגיות שאתן מתמודדת פלטפורמות מקוונות גדולות. בנוסף, יכול מודל זה לשמש, בהתאמות נחוצות, להתמודדות עם בעיות רגולציה שאתן מתמודדות חברות בתעשיות אחרות.