Opening the Black-Box: Improving Societal Understanding and Control Over Social Media Algorithms

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Alby Alex

Fall 2024

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

I. Introduction

Social media has slowly become the main method by which people disseminate and acquire information (Gottfried, 2024). People can simultaneously learn about new developments in a global war and catch up on what their friends are up to. However, the sheer amount of information available is overwhelming, so users unconsciously employ social media algorithms to curate and present only the posts that they will find most interesting (Pentina and Tarafadar, 2014). This works fine for posts from friends and colleagues, but it could produce significant negative effects when it is applied to news coverage. These effects are exacerbated for users whose only exposure to news is incidental contact through their social media feeds. As stated by Fletcher and Nielsen (2017, "Abstract"), "incidentally exposed users use significantly more online news sources than non-users".

Social media companies are intentionally very opaque about the effects that individual algorithm changes have on user dynamics and beliefs. They collect a massive amount of potentially useful data, but they keep it in house to avoid significant scrutiny (Andrejevic, 2013) and to maintain their advertising revenue. In a first of its' kind study, a group of academic researchers teamed up with Meta to determine how particular algorithm changes and behaviors influenced people's beliefs. The study was able to produce concrete results that critically examined the benefits and drawbacks of each proposed algorithmic solution. As stated by Garcia (2023a, p.40), an author that reviewed the results of the studies, "Meta and other companies must honestly and openly embrace regulated collaborations to scale up studies such as those now reported, and thereby to have a responsible role in our digital society". Although academics are more qualified than engineers to determine the social effects of algorithmic changes, this type of collaboration is unlikely to occur again without some type of systemic change.

This paper utilizes Geels' (2007) multilevel perspective to study previous examples of sociotechnical change in large systems to determine both effective actions and necessary circumstances for those actions to succeed. This analysis will compare this system to the labor movement in the United States, which was successful in changing the system and producing a new status quo with legally enforceable worker protections through both worker actions and effective legislation. Understanding the external circumstances and direct causes that led to these shifts will provide a roadmap to affecting change in the behavior of social media companies. The success of the labor movement required intervention from the federal government in addition to union actions. In this paper, I argue that user action must be timed perfectly along with structural pressures to successfully change the behavior of social media companies.

II. Problem Definition: Social Media Platforms Use their Structural Power to Maintain the Information-Limited Status Quo

The Users of Social Media Have Limited Control of What They See

Users of social media platforms have a very limited understanding of the algorithms that determine what they see and interact with, but the algorithms have almost unlimited leeway to collect and utilize any data that they can from the user (Reviglio & Agosti, 2020). Companies attempt to maintain this information imbalance in any way that they possibly can because collecting all this data is how they maintain their large advertising revenues. Any loss in collected data translates to billions in lost revenue. This was proven when Apple allowed users to opt out of ad identification and personalization in apps. Over the next year, Meta's expected revenues dropped by ten billion dollars (Newman, 2022). As such, social media platform owners have significant monetary incentives to keep collecting as much data as possible and to avoid disclosing how it is used.

This policy extends to algorithm changes, which affect both individual users as well as the news agencies that post their content on the platforms (Cornia et al., 2018). As stated by Devito et al. (2017, p.3163), "As algorithmically-driven content curation has become an increasingly common feature of social media platforms, user resistance to algorithmic change has become more frequent and visible". For news agencies, these changes may lead to their content performing worse in their aftermath. For example, if the algorithm starts to prioritize videos under 10 minutes, then any previously posted videos or planned videos will need to be re-edited to avoid a loss of engagement. As such, each time the algorithm changes, news agencies need to spend time and money on figuring out what changed and how it affects their past and future content. Altogether, this frustration and extra effort involved with each algorithm change leads all types of users to develop a dislike of any change, regardless of how necessary it is.

The best way for these users to avoid these negative consequences from changes is to achieve 'algorithmic sovereignty', or a scenario where users can review exactly what an algorithm is doing with the data that it receives from the user (Reviglio & Agosti, 2020). This would simultaneously give users a list of what data the platforms collect from them and a roadmap to how the platform utilizes it. However, the biggest difference would be in how users feel about changes to the platform. If the algorithm is public, the company that owns the platform can explain why they implemented a particular change. As a result, users might understand why their feeds changed and be more forgiving of the platform (Devito et al., 2017). Social media platforms are environments where millions of people interact daily, and the algorithms that govern them need to be updated to respond to new threats. Making the algorithms public would give users a clear idea of why platforms are taking the actions that they are.

Researchers Attempt to Find Correlations with the Limited Resources They Have

Academic research on social media algorithms is currently conducted using very scientifically unsound methods due to restrictions on how much control any user has over their own algorithm. In one case, researchers trying to study the effects of politically biased feeds generated their feeds by making study subjects follow bots that posted partisan content, as shown in Figure 1 (Bail et al., 2018). While this seems to be a reasonable solution for this study, it does not mirror what the average person would interact with through their own social media feeds. A purely partisan feed is highly unlikely; instead, individual users are more likely to encounter political posts interspersed with personal posts. The so-called treatments were not exhaustive, limiting their applicability to actual users of social media. Researchers have no real mechanisms to emulate this behavior, leading them to do the best they can.

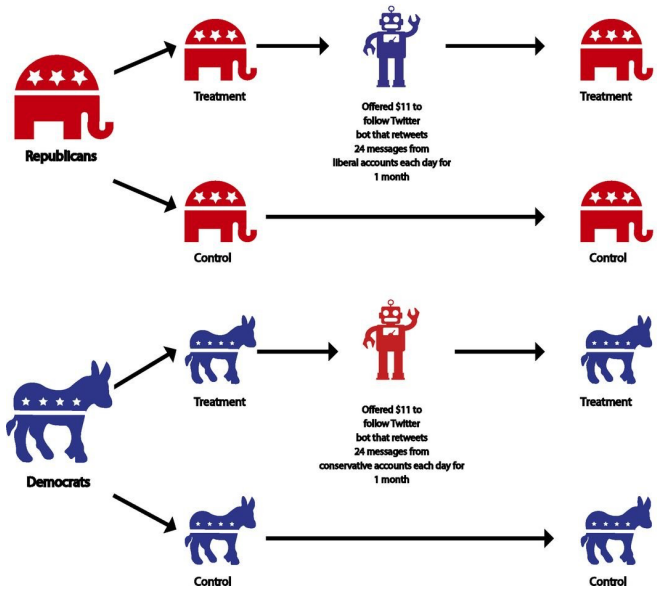

Figure 1. Experimental Design for a Political Polarization Study. A diagram of the design that one study had to undertake because of a lack of control over algorithms or controlled social media feeds (adapted from Bail et al., 2018, p.9218)

This issue also extends to the data collection process. In a review of research on political polarization caused by social media, Kubin et al. (2021, p.198) state, “our quantitative analyses

of the content of media highlighted an intense focus on analyzing Twitter. While this is likely due to the ease at which researchers can scrape data from Twitter as compared to other social networking sites – it makes it difficult to understand whether similar trends occur on other social media platforms." Unfortunately, this means that studies on the more popular platforms, like Facebook or Instagram, are limited. Current research on social media algorithms is hamstrung by the platforms, and it needs to change to develop a full understanding of what the algorithms do to personal dynamics.

If researchers are given access to the data that social media companies already collect and the details of the algorithms, they can test numerous hypotheses and conduct significant research (Garcia, 2023). This was proven by a group of joint studies that Meta conducted with a group of researchers that were given full authority to alter algorithms for a subset of users and monitor their behavior afterwards. The studies focused on understanding how particular algorithmic changes could help slow down partisan polarization. While one individual study was unable to find a working approach, the work that the researchers did help Meta engineers by outlining what approaches would not be sufficient (Guess et al., 2023). A review of these studies calls for more of these collaborative studies to understand other social dynamics that are exacerbated by social media (Garcia, 2023). These researchers were given the proper tools they need to conduct their research, and they provided results that affected that meta was planning to go in. Engineers do not have adequate expertise to understand the social effects of specific algorithmic behaviors, so this role must be assumed by academics who are well-versed in the domain.

Gap in Knowledge on How to Implement Changes

Other engineering disciplines have evolved to be more transparent about their methods independently, without direct external pressure; a notable example is how civil engineers adopted

project peer reviews (PPR) (Gustafson, 1990). As described by Preziosi (1988, p.46), "[American Society of Civil Engineers] … defines PPR as 'structured, comprehensive and thorough fact-finding process conducted by one or more senior professionals who are separate and independent from the organization preparing a project design'." PPR was adopted as a response to a 1981 collapsing walkway in Kansas City and a 1987 collapsing slab in Bridgeport. Together, both these events led to more than 100 deaths, and civil engineers realized that they had to change something to avoid a negative reputation for their profession. Their response was effective, as evidenced by a peer review of a dam in the late 1970s that removed numerous design inconsistencies and led to the use of alternative construction methods that saved time and money on the project. Although this adoption of a similar program was a success, this is unlikely to occur with social media platforms due to the lack of a disaster that is directly caused by social media.

Social media companies have no incentive to change on their own because they are unlikely to face a scenario where they must save face like civil engineers did. Civil engineers build physical objects that are used by many people and whose failures can directly result in death or disability. On the other hand, even though social media platforms interact with millions of people daily, they alone are very unlikely to directly cause deaths. The negative effects of algorithmic decisions affect users though gradual progressions that are difficult to track. As a result, social media companies can easily avoid allegations of direct misconduct that leads to disaster. This conclusion establishes that the only way in which these platforms adopt anything close to "algorithmic sovereignty" is if users pressure them. However, this process is not direct or simple, and it requires a particular set of circumstances and a set of specific actions by users to be effective.

This paper uses Geels' (2007) multilevel perspective to establish both the necessary circumstances and prudent actions for user driven change to be effective in convincing social media companies to adopt "algorithmic sovereignty". This process will be compared to specific case studies of the labor movement in the United States that improved workers' rights through actions undertaken by workers themselves. However, Geels' multilevel perspective also helps to understand which large-scale circumstances were necessary for this change to be successful. Similar to how workers are vital for producing revenue for bosses, users on social media are the product that the companies sell to generate revenue. Without users, there is no ad revenue from companies that want to display their products to millions of people all around the world. The bulk of what makes social media interesting are the other users, which gives them a level of power over social media companies that is like what workers had over their own bosses.

III. Research Approach: Utilizing Geels' Method to Determine Potential Next Steps for Social Media Users

Geels' (2007) approach simultaneously provides reasoning for why sociotechnical changes take time to occur and why the multilevel approach provides the best explanation for this change. The three levels that Geels outlines are the niche level, regime level, and the landscape level. The niche level represents the micro level; it is where the actions of individual actors and firms are considered. The regime level represents the existing system of rules and agreements that currently drive development and progress in the sociotechnical system. In Geels' words, "regime rules account for the stability and momentum of existing sociotechnical systems" (2007, p.128). Regimes are what maintain the momentum that was created at the establishment of the system. The landscape level represents the macro level of the model, corresponding to major changes in environment that the system operates within. Examples given by Geels include

the “material and spatial arrangements of cities, factories, and electricity infrastructures” (2007, p.129). These different levels must each align with each other in a specific fashion to produce lasting change in the sociotechnical system.

To demonstrate how these levels interact to produce change, Geels dives into the development of changes in the Dutch highway system from 1950 to 1990. He contrasts how the two separate periods of change had different contributing factors that led to one success and one failure. Geels starts his analysis of the highway system by first finding positive developments at the niche level that are gradually adopted by other actors at that level until they stabilize in a “dominant design”. Next, he looks for a landscape development that creates “windows of opportunity for novelties”. As Geels outlines, once this development and a dominant design intersect, the “new configuration breaks through…Adjustments occur in the socio-technical regime” (2007, p.130). This process is illustrated in Figure 2 below and provides a clear roadmap for how Geels’ method can be applied to the study of any sociotechnical change.

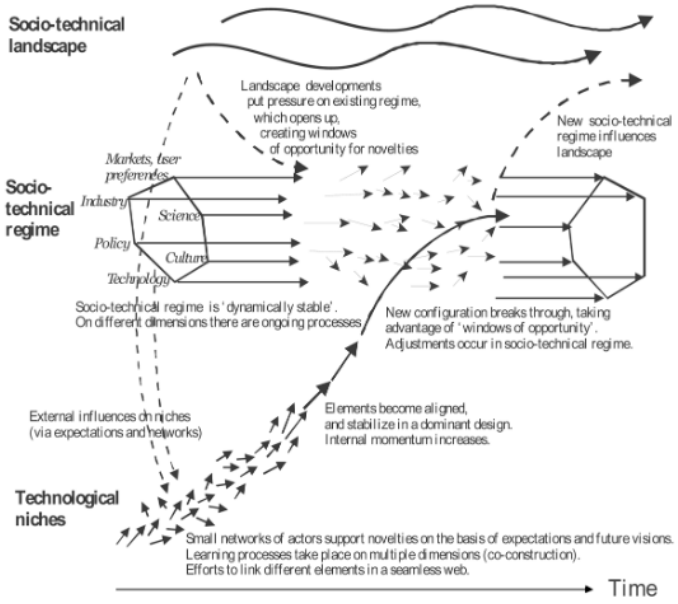

Figure 2. Sociotechnical System Changes with the Multi-Level Perspective. An illustration of how the multi-level perspective can be used to trace changes in sociotechnical systems (adapted from Geels, 2007, p.130).

This paper analyzes review articles that recount the history of the United States labor movements to understand what circumstances and what actions are necessary for a bottom-up change in a sociotechnical system. During this analysis, I first worked to determine the niche developments that eventually became adopted at large. This corresponds to the creation of unions or specific collective actions undertaken by groups of workers. Next, I took time to understand how this action propagated among different groups of workers until it became more widespread and analyzed how successful it was. If the action failed to propagate, I looked for a reason why. However, if it succeeded, I looked for a landscape level event that enabled this niche development to expand its influence and application and break through to the regime. In the case of the labor movements, this generally corresponded to worker-friendly legislation. Finally, I determined the net effect of the changes on the overall regime that the system has. In other words, which protections for workers became standard.

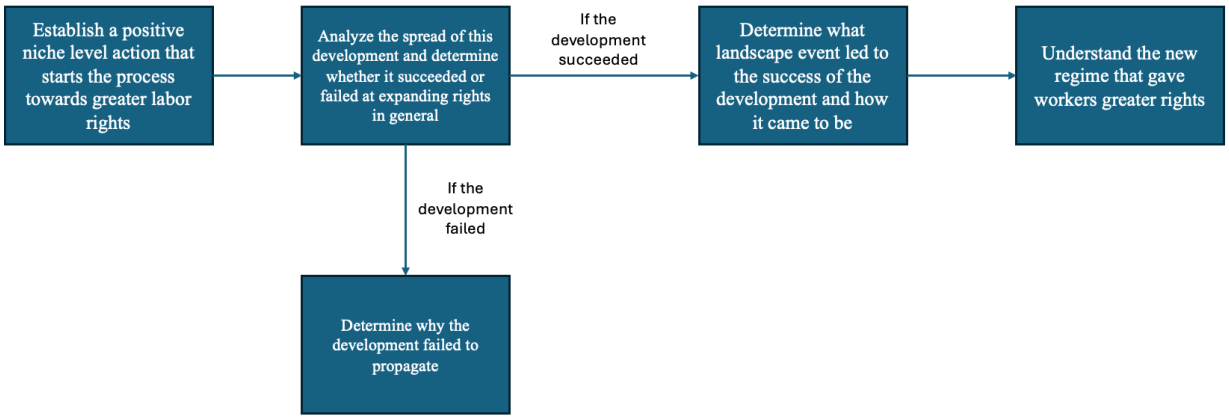

Figure 3. Analyzing Labor Actions. Both successful and unsuccessful movements for sociotechnical change during the labor movement can be analyzed using Geels' method (Created by author).

Geels' approach holds promise for evaluating labor movements because historical records tend to focus exclusively on the collective action at the niche level rather than considering the necessary landscape developments that led to successful changes (Compa, 2014). In addition,

Geels' process starts from the niche level and works its way up, mirroring the type of change that I believe is necessary to change social media platforms. However, Geels' research has one limitation: Geels only applies the method to a physical engineering task that is easy to monitor for progress and change. The sociotechnical change that I am proposing is more abstract and will require a different definition of progress. As such, I chose to apply Geels' method to the labor movements because they provide examples of change that does not have associated physical evidence.

Labor movements provide an effective proxy for change in social media platforms because users on social media have a similar relationship with social media companies as workers had with their bosses. Most social media platforms pay high-engagement users to convince them to continue creating content (Tang et al., 2012). This money comes in the form of a cut of the ad revenue that the user generates. As established before, social media companies do not like giving up ad revenue. However, the companies understand that these power users are what keep most silent users on the platform. As a result, power users have a unique form of leverage over these companies. If they decide to stop creating, the companies will have no choice but to listen. In addition, power users hold sway over their audiences and can convince them to adopt their stances. This collective effect produces an environment where social media companies are beholden to these power users just as bosses were beholden to their workers. As such, analyzing the labor movements will provide an effective list of actions and circumstances for implementing change in social media.

IV. Results: Success only happens with the right moment and right circumstances

Geels' (2007) method outlines a progression that needs to occur in a sociotechnical system to produce lasting change. Throughout my research process, I attempted to find the

corresponding steps in the labor movement that represented each of Geels' milestones. This section of the paper is designed as a narrative retelling of what happened in the labor rights movement. The narrative is split into three parts, each followed by a paragraph that connects that part to Geels' framework.

Collective bargaining has existed in the US since at least 1806 (Compa, 2014). However, these actions did not find any lasting success for much of the 19th century due to various judicial rulings. One of the notable cases was the Philadelphia Cordwainers case in 1806, where the court ruled that skilled shoemakers' plan to uniformly increase prices was an "unnatural" means of fixing their salary. Courts throughout the 19th century treated collective bargaining as a criminal conspiracy, and this allowed bosses to crush movements as they found necessary. As stated by Adler (2006, p.312), "employers used all available methods to crush these fledgling unions including the full weight of a hostile legal system". Even when courts did not use the criminal conspiracy theory, they constantly shut down labor actions because of "supposed potential for violence". Furthermore, as stated by Compa (2014, p.92), "they allowed employers to require newly hired employees to sign a contract promising never to join a union. Even where they did not impose such a requirement, employers were free to dismiss workers if they joined a union and wanted to bargain collectively." Even when workers were successful in striking, Compa (2014, p.92) states, "Police and National Guard forces often violently suppressed them. So did private security forces hired by employers, such as the Pinkerton Detective Agency".

The anti-labor stances of the state courts led to repeated failures for labor actions. In Geels' framework, this corresponds to niche changes that do not propagate to the regime due to a lack of landscape changes. Unfocused, random boycotts of social media will result in the same type of failures. The only way to organize these types of boycotts will be through the same social

media platforms that users are trying to change. Companies will not be able to use the legal system to defend their platforms, but they will be able to quash these movements through bans and algorithmic de-ranking. As a result, the chosen actions must be focused and targeted with a clear plan of action and appropriate circumstances to produce lasting change.

Labor movements pushed for change by targeting the ownership of individual companies for years with little success due to structural barriers like state courts and individual barriers like Pinkertons. These failures cumulated in an eventual change in strategy by unions. As stated by Compa (2014, p.92), "American society began looking to federal legislation to address continuing labor conflict and to develop a unified national policy with regard to collective bargaining in the private sector." The first meaningful pro-worker legislation passed by Congress was the Railway Labor Act in 1926 (Compa, 2014). This legislation established a system for unions and employers to debate their collective bargaining agreements. During this arbitration period, workers were protected from losing their employment, as the terms of their previous contract remained in action. Congress chose to address railroad workers first because of how devastating a threatened strike in this industry would be to the country's economy. As a result, workers in other industries asked for the same protections given to railway workers. Within the next decade, Congress packed the National Labor Relations Act to cover most other private sector workers. State courts stymied the progress of the labor movement throughout the 19th century, but Congress produced proper legislation in the 20th century to avoid more dangerous strikes.

After a history of repeated failure and slight successes, the labor movement achieved long-term benefits only after the federal government stepped in. In the terms of Geels' framework, this legislation was the result of the landscape-level change that was needed for

successful regime change. The corresponding landscape change was the country's economic dependence on railroads, which justified Congressional fears of the negative effects of a railroad strike. This shines a light on the circumstances that are necessary to change the behavior of social media companies. Congress needs to step into the dispute, or companies will pull any trick necessary to maintain control of their userbase. State courts will help them in this process because of their economic interests. The federal government has the power to take on these companies and can be influenced to action by some type of consequence, be it public opinion or economic decline.

Once the labor movement understood the necessity of federal response, they rerouted their efforts towards building influence in Congress. This change in priorities runs contrary to the notion that the labor movement was an apolitical movement that focused only focused on economic pressure. As Kimeldorf (1992, p.496) states, "recent historical research has shown that organized labor in the United States has never fully lived up to its apolitical image". While the Railway Labor Act was passed to avoid crippling labor action, the subsequent legislation was constructed in tandem with unions. As stated by Kimeldorf (1992, p.496), the national American Federation of Labor "forged temporary ties with the Democratic Party in order to pass various legislation." This partnership allowed them to work together to produce the National Labor Relations Act.

Sweeping federal legislation was the biggest driver behind the labor movement's successes, and it was only possible because of collaboration with politicians. In the terms of Geels' framework, labor organizers aided and abetted the process of regime alteration through their work with legislators. They were able to produce a new regime that had a place for their concerns and values. Once the federal government decides to get involved, users need to partner

with legislators to create laws that direct and incentivize social media companies to be responsive to their users. This is the only viable path towards achieving "algorithmic sovereignty" on social media platforms.

The labor movement presents a clear view of what events needed to occur to produce lasting change in a large system. First, laborers would group together in their workplaces to form unions. Then, the unions had two main options. The first option was to engage in collective action, like strikes and walkouts. Unfortunately, these actions led to either legal union dismantlement by state courts or forceful dismantlement by police or Pinkertons. The other option was to just threaten collective action until Congress decided to step in to protect a particular industry. This option was exercised by railroad worker unions, and it led to the eventual success of the labor movement. The results described above are summarized in Figure 4 below. The labor movement was not part of a sociotechnical system, as there was no real technical component. However, applying Geels' (2007) multilevel perspective to the labor movement reveals that the process of change in any large system mirrors change in sociotechnical systems. Small, user-driven actions are just niche level events that have no staying power unless they are allowed to permeate to the regime by a landscape level event. As such, any actions taken by social media users today is unlikely to produce any meaningful results until something influences Congress to take an interest in what social media platforms are doing.

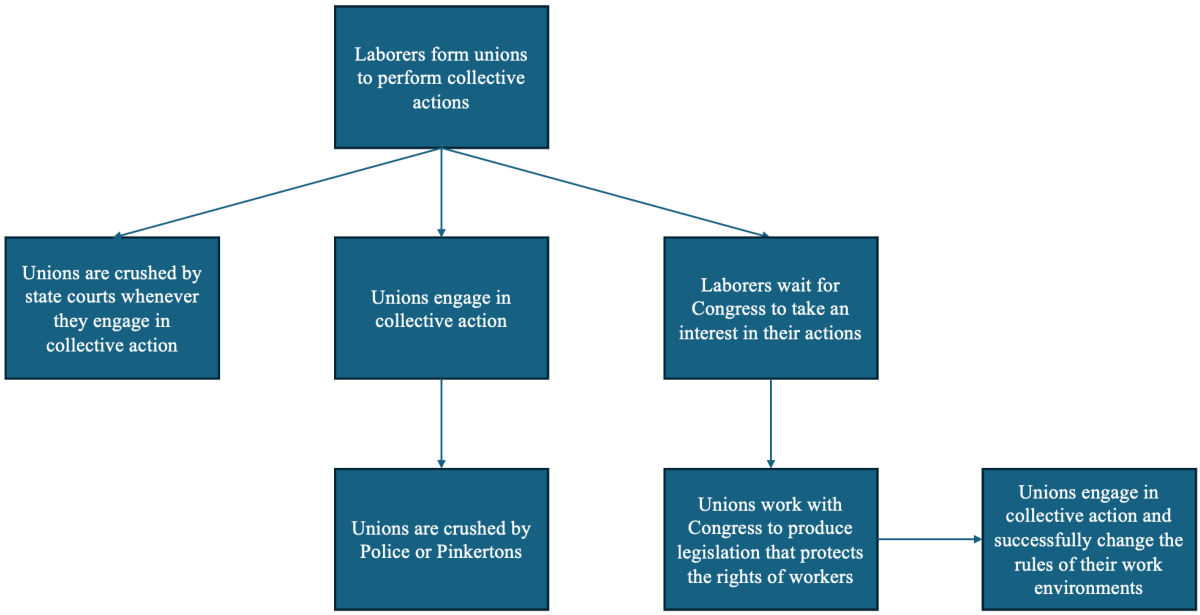

Figure 4. A Diagram of Potential Outcomes from Unionization Actions during the Labor Movement. The only successful path for unions is to work with Congress. (Created by author)

V. Conclusion

User action alone cannot cause social media companies to adopt algorithmic transparency. The companies themselves will not change their behavior on their own, as engineers only change their sociotechnical systems towards transparency when well-publicized, catastrophic events occur. The prominent historical example for user-driven change is the labor movement of the 19th and 20th century in the United States. As evidenced by this movement, these user-driven actions need to involve the government to produce tangible results. However, this government involvement only occurred because of specific circumstances. As Geels' (2007) multilevel perspective outlines, changes at the niche level cannot progress to the regime level without some type of landscape level pressure. In the case of the labor movement, this pressure came from possible economic effects of a railroad labor strike. Users of social media do not have any sort of similar influence, which limits the possible pressure that these users can exert on Congress. While this limits the current options for user-driven change in social media, it does not

change the idea that users should push for change by demanding more stringent regulation instead of trying to negotiate with large corporations. Social media is the primary form of large-scale communication and news distribution for a significant portion of the population, and it is positioned to take an even larger role as the people who grew up with it get older. As this usage grows, users may acquire enough influence to pressure Congress to act. When this moment presents itself, users must take advantage to ensure that social media companies adapt algorithmic sovereignty for their platforms.

References

Adler, J. (2006). The past as prologue? *Public Personnel Management, 35*(4), 311–329. https://doi.org/10.1177/009102600603500405

Andrejevic, M. (2013). Public service media utilities: Rethinking search engines and social networking as public goods. *Media International Australia, 146*(1), 123–132. https://doi.org/10.1177/1329878x1314600116

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences, 115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Boczkowski, P., Mitchelstein, E., & Matassi, M. (2017). Incidental news: How young people consume news on social media. *Proceedings of the Annual Hawaii International Conference on System Sciences*. https://doi.org/10.24251/hicss.2017.217

Compa, L. (2014). An overview of collective bargaining in the United States. In *El derecho a la negociación colectiva* (pp. 91–98). Consejo Andalus de Relaciones Laborales.

Cornia, A., Sehl, A., Levy, D., & Nielsen, R. (2018). *Private sector news, social media distribution, and algorithm change.* Reuters Institute for the Study of Journalism.

DeVito, M. A., Gergle, D., & Birnholtz, J. (2017). "Algorithms ruin everything": #RIPTwitter, folk theories, and resistance to algorithmic change in social media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3025453.3025659

Fletcher, R., & Nielsen, R. K. (2017). Are people incidentally exposed to news on social media? A comparative analysis. *New Media & Society*, *20*(7), 2450–2468. https://doi.org/10.1177/1461444817724170

Garcia, D. (2023a). Influence of facebook algorithms on political polarization tested. *Nature, 620*(7972), 39–41. https://doi.org/10.1038/d41586-023-02325-x

Geels, F. (2007). Transformations of large technical systems: A multilevel analysis of the Dutch highway system (1959-2000). *Science, Technology, and Human Values 32*(2), 123-149.

Gottfried, J. (2024, January 31). *Americans' social media use*. Pew Research Center. https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., … Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an

election campaign? *Science, 381*(6656), 398–404. https://doi.org/10.1126/science.abp9364

Gustafson, G. D. (1990). Why project peer review? *Journal of Management in Engineering, 6*(3), 350–354. https://doi.org/10.1061/(asce)9742-597x(1990)6:3(350)

Kimeldorf, H., & Stepan-Norris, J. (1992). Historical Studies of Labor Movements in the United States. *Annual Review of Sociology*, *18*, 495–517. http://www.jstor.org/stable/2083464

Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association, 45*(3), 188–206. https://doi.org/10.1080/23808985.2021.1976070

Newman, D. (2022, October 12). *Apple, Meta and the $10 billion impact of privacy changes*. Forbes. https://www.forbes.com/sites/danielnewman/2022/02/10/apple-meta-and-the-ten-billion-dollar-impact-of-privacy-changes/ [magazine article]

Pentina, I., & Tarafdar, M. (2014). From "information" to "knowing": Exploring the role of social media in contemporary news consumption. *Computers in Human Behavior, 35*, 211–223. https://doi.org/10.1016/j.chb.2014.02.045

Preziosi, D. (1988). Reviewing Peer Review. *Civil Engineering*, *58*(11), 46. https://proxy1.library.virginia.edu/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fscholarly-journals%2Freviewing-peer-review%2Fdocview%2F228531568%2Fse-2%3Faccountid%3D14678

Reviglio, U., & Agosti, C. (2020a). Thinking outside the black-box: The case for "algorithmic sovereignty" in social media. *Social Media + Society*, *6*(2). https://doi.org/10.1177/2056305120915613

Tang, Q., Gu, B., & Whinston, A. B. (2012). Content contribution for revenue sharing and reputation in social media: A dynamic structural model. *Journal of Management Information Systems, 29*(2), 41–76. https://doi.org/10.2753/MIS0742-1222290203