# Sunsetting "Shadowbanning"

*Gabriel Nicholas*

*"The radical left tries to blacklist you on social media .. . . On Twitter it's very hard to come on to my account. It's very hard to get out a message . . . . If you're a conservative, if you're a Republican, if you have a really big voice, I guess they call it shadow ban. Shadow ban. They shadow ban you and it should be illegal."*
- President Donald Trump, "Save America Rally", January 6, 2021[1]

*"I know this sounds kind of tinfoil hatty but . . . when you get a post taken down or a story, you can set a timer on your phone for two weeks to the godd\*mn f\*cking minute and when that timer goes off you'll see an immediate change in your engagement. They put you back on the Explore page and you start getting followers."*
- Anonymous Instagram user[2]

## introduction

On April 27, 2018, sex worker and digital rights activist Jessie Sage attended "Theorizing the Web," an interdisciplinary conference about the internet and society.[3] She was there, according to my interview with her,[4] to talk about FOSTA-SESTA, a bill President Trump had signed into law three weeks earlier that limits free speech protections for online services that fail to curb sex trafficking. Sage saw the effects of the law

---

[1] *Donald Trump Speech "Save America" Rally Transcript*, Rᴇᴠ (Jan. 6, 2021), https://www.rev.com/blog/transcripts/donald-trump-speech-save-america-rally-transcript-january-6.

[2] Josh Constine, *Instagram Now Demotes Vaguely 'Inappropriate Content'*, TᴇᴄʜCʀᴜɴᴄʜ (Apr. 10, 2019, 4:38 PM), https://social.techcrunch.com/2019/04/10/instagram-borderline.

[3] *Theorizing the Web*, https://www.theorizingtheweb.org (last visited Jan. 27, 2023).

[4] Zoom Interview with Jessie Sage, Co-Founder, Peepshow Media (Mar. 17, 2022).

immediately: Craigslist Personals and Backpage shut down only days after the bill's passage, and other online infrastructure that sex workers depended on for safely finding work were under threat.[5] "Theorizing the Web" seemed to Sage like a natural forum to discuss FOSTA-SESTA issues. It had panels on queer internet culture, online representations of bodies, and one paper talk called "Recuperating the Cybernetic Libido." To her surprise though, conference attendees hardly mentioned the bill. So, she did what one does when they want to shift The Discourse—she went onto Twitter.

Sage tweeted about FOSTA-SESTA under the conference and panel hashtags,[6] but despite a small, active, and generally supportive conversation happening alongside the event online, her tweets got no likes or replies from other attendees. Everything appeared otherwise normal on Sage's device, but she had experienced Twitter quietly flagging her content before, so she checked the conference hashtags from someone else's phone. Sure enough, her posts did not appear.

Sage's story illustrates the unnerving nature and possible dangers of shadowbanning. However, I want to use it here to argue a more discursive point: that the word "shadowbanning" itself precludes productive conversation about undisclosed content moderation and therefore should be retired. In another essay in this series, Gillespie discusses how "shadowbanning" has become a catchall term for any hidden action a platform takes against its users and thus reflects the outdated concept of platforms only removing content instead of reducing its visibility.[7] In this essay, I take Gillespie's point one step further and argue that the vagueness, politicization, and conspiratorial implications of the word "shadowbanning" focus public discourse on epistemically unresolvable arguments over whether platforms secretly silence certain individuals, instead of more productive debate over when moderating a user's content without informing them is or is not justified. The change is subtle, but by adopting a better defined, less pejorative term such as "uninformed

---

[5] *See* Danielle Blunt & Ariel Wolf, *Erased: The Impact of FOSTA-SESTA*, HACKING//HUSTLING (Jan. 2020), https://hackinghustling.org/wp-content/uploads/2020/01/HackingHustling-Erased.pdf.

[6] Jessie Sage (@sapiotextual), Twitter (Apr. 27, 2018, 1:05 PM), https://twitter.com/sapiotextual/status/989913464746409984; Jessie Sage (@sapiotextual), Twitter (Apr. 28, 2018, 8:13 PM), https://twitter.com/sapiotextual/status/990383512879591425.

[7] Tarleton Gillespie, *Reduction / Borderline content / Shadowbanning*, Platform Governance Terminologies Essay Series, YALE INFO. SOC'Y PROJECT & WIKIMEDIA INITIATIVE ON INTERMEDIARIES & INFO. 8-9 (July 20, 2022), https://law.yale.edu/sites/default/files/area/center/isp/documents/reduction_ispessayseries_jul2022.pdf.

content moderation," the public can shift away from conspiratorial thinking; begin to acknowledge situations where so-called "shadowbanning" may be justified, such as stopping harassers from creating new accounts after being blocked; and most importantly, be better equipped to push platforms to reveal their content moderation policies, especially when they do not inform users of moderation.

Sage's story and the question it begs—was she shadowbanned? —elucidate the pitfalls of existing language. The problems start with definitions. For most of the word's fifteen-plus year history,[8] people used "shadowbanning" to refer to instances in which a social media service made a user's posts invisible to everyone but the poster themselves.[9] Online community administrators and moderators could shadowban misbehaving users to protect other community members from abuse without tipping off spammers and trolls. As social media services evolved and grew to industrial scales, what people consider "shadowbanning" broadened to encompass new methods of opaque content moderation. Users now use the word "shadowban" to describe a wide range of actions, including social media services suppressing usernames in search, deprioritizing content in algorithmically curated feeds, and, in Sage's case, hiding posts under hashtags.[10]

Today, "shadowbanning" describes both a practice and a political critique of said practice. The word has become shorthand for social media censorship, and not just for sex workers: conservatives,[11] Black Lives

---

[8] *See* Dieter Bohn, *One of Twitter's New Anti-Abuse Measures Is the Oldest Trick in the Forum Moderation Book*, THE VERGE (Feb. 16, 2017), https://www.theverge.com/2017/2/16/14635030/twitter-shadow-ban-moderation.

[9] *See* Samantha Cole, *Where Did the Concept of 'Shadow Banning' Come From?*, VICE (July 31, 2018), https://www.vice.com/en/article/a3q744/where-did-shadow-banning-come-from-trump-republicans-shadowbanned.

[10] *See* Danielle Blunt, Emily Coombes, Shanelle Mullin & Ariel Wolfe, *Posting Into the Void: Studying the Impact of Shadowbanning on Sex Workers and Activists*, HACKING//HUSTLING (2020), https://hackinghustling.org/posting-into-the-void-content-moderation.

[11] *See* Ashley Rae Goldenberg & Dan Gainor, *Censored! How Online Media Companies Are Suppressing Conservative Speech*, MEDIA RSCH. CTR., THE HEARTLAND INST. (Apr. 19, 2019), https://www.heartland.org/_template-assets/documents/censored-howonlinemediacompaniesaresuppressingconservativespeech-digitalcopy.pdf.

Matter activists,[12] plus-sized influencers,[13] trans folks,[14] and many others[15] use it to describe their experiences of feeling marginalized by disproportionate content moderation online. However, exogenously determining whether an individual is really shadowbanned (not to mention evaluating more macro trends) is nearly impossible, since from the end users' perspective, shadowbanning often looks identical to a technical glitch or a post being simply unpopular. But perceived shadowbanning is widespread—according to a survey from the Center for Democracy & Technology, nearly one out of ten U.S. social media users believe that they have, at some point, been shadowbanned.[16]

In this essay, I first give a history of the term "shadowbanning" and explain how it reached its current, untenable state. I then describe three failures of the term: first, its vagueness allows social media companies to avoid adequately describing their moderation practices; second, its unprovable nature allows political actors to appropriate the term to their own ends; and third, its conspiratorial implications prevent social media companies from using the term at all, pushing their moderation practices further into the dark. Finally, I recommend that platforms disclose their policies for when they do or do not inform a user that their content has been moderated and use that disclosure as a chance to shift to a better

---

[12] *See* Meira Gebel, *Black Creators Say TikTok Still Secretly Hides Their Content*, DIGIT. TRENDS (July 21, 2020), https://www.digitaltrends.com/social-media/black-creators-claim-tiktok-still-secretly-blocking-content.

[13] *See* Chanté Joseph, *Instagram's Murky 'Shadow Bans' Just Serve to Censor Marginalized Communities*, THE GUARDIAN (Nov. 8, 2019), http://www.theguardian.com/commentisfree/2019/nov/08/instagram-shadow-bans-marginalised-communities-queer-plus-sized-bodies-sexually-suggestive.

[14] *See* Imara Jones, *Shadowbanning on Facebook: As a Black Transgender Woman, Why Am I Invisible?*, FAST CO. (Feb. 23, 2022), https://www.fastcompany.com/90721395/shadowbanning-on-facebook-as-a-black-transgender-woman-why-am-i-invisible; *Exclusive: An Investigation into Algorithmic Bias in Content Policing on Instagram*, SALTY (Jan. 23, 2021), https://www.saltyworld.net/algorithmicbiasreport-2.

[15] *See, e.g.*, *The Attacks on Palestinian Digital Rights*, 7AMLEH – THE ARAB CTR. FOR THE ADVANCEMENT OF SOC. MEDIA (May 21, 2021), https://7amleh.org//storage/The%20Attacks%20on%20Palestinian%20Digital%20Rights.pdf; Carolina Are, *The Shadowban Cycle: An Autoethnography of Pole Dancing, Nudity, and Censorship on Instagram*, FEMINIST MEDIA STUD. (May 19, 2021), https://doi.org/10.1080/14680777.2021.1928259; Blunt et al., *supra* note 10; Fergus Ryan, Audrey Fritz & Daria Impiombato, *TikTok and WeChat: Curating and Controlling Global Information Flows*, AUSTL. STRATEGIC POL'Y INST. 72 (Sept. 2020), https://s3-ap-southeast-2.amazonaws.com/ad-aspi/2020-09/TikTok%20and%20WeChat.pdf.

[16] Gabriel Nicholas, *Shedding Light on Shadowbanning*, CTR. FOR DEMOCRACY & TECH. (Apr. 26, 2022), https://cdt.org/wp-content/uploads/2022/04/remediated-final-shadowbanning-final-050322-upd-ref.pdf.

defined, less appropriable term, such as "undisclosed content moderation."

### 1.    A History of "Shadowbanning"

The origins of shadowbanning—also known as hell banning, ghost banning, or stealth banning, each with and without spaces[17]—can be traced back to online bulletin boards, the earliest social networks from the days when the web was still a series of connected "intranets," housed mostly in universities and government agencies.[18] On Citadel BBS, for example, an online bulletin board system from the 1980s, administrators and moderators could flip what was called a "coventry bit,"[19] a switch that would make an individual post or all of a given user's posts only visible to that user.[20] The Whole Earth 'Lectronic Link usenet (the WELL), one of the most influential virtual communities of the 1990s, similarly had a "bozo filter" to silently hide the activity of trollish "netbozos." In 1993, *Wired* writer Howard Rheingold wrote, "I'm generally skeptical of technical solutions to social problems, but bozo filters are evidence that tools sometimes work better than rules."[21]

When intranets were phased out in favor of the unified "internet," burgeoning social networks used shadowbanning to manage the newfound scale of bad actors they had to manage. As founder of StackOverflow Jeff Atwood wrote, "I've personally talked to people in charge of large online communities – ones you probably participate in every day – and part of the reason those communities haven't broken down into utter chaos by now is because they secretly hellban and slowban their most problematic users."[22] Shadowbanning also became more automated over time. For example, Disqus, a comment-hosting service that was popular in the early 2010s, automatically hid a user's comments from everyone but the user herself if those comments received enough negative ratings.[23]

---

[17] *See* Jeff Atwood, *Suspension, Ban or Hellban?*, Coding Horror (June 4, 2011), https://blog.codinghorror.com/suspension-ban-or-hellban.

[18] *See* Cole, *supra* note 9.

[19] The term derives from the British phrase, "to send one to coventry," meaning give them the silent treatment.

[20] *See id.*; Bohn, *supra* note 8.

[21] Howard Rheingold, *Bozo Filters*, Wired (Jan. 1, 1993), https://www.wired.com/1993/01/bozo-filters.

[22] Atwood, *supra* note 17.

[23] *See Clive Thompson on the Taming of Comment Trolls*, Wired (Mar. 23, 2009), https://www.wired.com/2009/03/st-thompson-12.

Like any internet neologism, its exact origin is hard to trace,[24] but the word "shadowban" seems to emerge on Reddit around 2008.[25] For many years, shadowbanning was the only content moderation action that administrators on Reddit could take against abusive users. If a Reddit employee removed a user's comment or post, it still appeared to be unaltered and publicly viewable to the original poster. Reddit was open about the practice, but it was deeply unpopular with users. In a blog post entitled "On shadowbans," a Reddit administrator wrote, "We know it sucks. It sucks hard. It is awful and sneaky and completely our fault that it is still being used to punish normal users."[26] Reddit largely phased out shadowbanning in 2015 but saved it for occasional use in dealing with bots and spam rings.[27]

Large social media services—or at least Facebook—likely began to quietly practice shadowbanning sometime in the 2010s, as indicated by a handful of patents. The first such patent, filed November 30, 2011, was entitled "Moderating content in an online forum" and stated that "the social networking system may display the blocked content to the commenting user such that the commenting user is not made aware that his or her comment was blocked, thereby providing fewer incentives to the commenting user to spam the page or attempt to circumvent the social networking system filters".[28] Facebook filed another patent in 2015 for a system that automatically demoted potential disinformation content based on how users responded to it or whether certain links or keywords were present[29], and a third patent, that actually used the word "shadow ban", to mitigate spam posts on Facebook Marketplace.[30].

---

[24] It likely first appeared on the *Something Awful* forums. *See* Samantha Cole, *Where Did the Concept of 'Shadow Banning' Come From?*, VICE (Jul. 31, 2018), https://www.vice.com/en/article/a3q744/where-did-shadow-banning-come-from-trump-republicans-shadowbanned.

[25] This is when the r/shadowbanned subreddit appeared for users to discuss their experiences being shadowbanned on the website.

[26] krispykrackers, *On Shadowbans*, REDDIT (July 28, 2015), www.reddit.com/r/self/comments/3ey0fv/on_shadowbans.

[27] *See* powerlanguage, *Account Suspensions: A Transparent Alternative to Shadowbans*, Reddit (Nov. 10, 2015), www.reddit.com/r/announcements/comments/3sbrro/account_suspensions_a_transparent_alternative_to.

[28] U.S. Patent No. US20150163184A1 (Jul. 16, 2019), https://patents.google.com/patent/US20150163184A1.

[29] U.S. Patent No. US20190171681A1 (Jul. 5, 2022), https://patents.google.com/patent/US20190171681A1.

[30] U.S. Patent No. US11283743B1 (Mar. 3, 2022), https://patents.google.com/patent/US11283743B1.

Shadowbanning as both a term and concept came to the attention of the mainstream public in 2018, when a *Vice* article found evidence that Twitter was not autofilling the handles of many prominent Republicans in its search bar, including Republican National Committee chair Ronna McDaniel, Chief of Staff Mark Meadows, and Representatives Jim Jordan and Matt Gaetz.[31] In a post-mortem blog post, Twitter said it had rolled out an update to improve the health of conversation online, but the change had accidentally affected hundreds of thousands of accounts across the political spectrum.[32] Twitter fixed the bug, but not before the debacle sparked widespread anger from conservatives accusing Twitter of bias in its content moderation practices[33] and retaliatory eye rolling from liberals.[34]

The 2018 Twitter incident kicked off a trend of many different groups using the word "shadowbanning to describe a wide array of undisclosed content moderation experiences, not just surreptitious removal. One of the first comprehensive frameworks describing how users experience shadowbanning came from the sex worker advocacy and research collective *Hacking // Hustling* in their report, "Posting into the Void."[35] Black Lives Matter activists accused TikTok of shadowbanning their videos by not including them in other users' "For You" pages.[36] And social media scholar Carolina Are wrote an autoethnography of how pole dancers deal with shadowbanning on Instagram, which included different ways of decreasing the discoverability of their content.[37]

---

[31] *See* Alex Thompson, *Twitter Appears to Have Fixed "Shadow Ban" of Prominent Republicans Like the RNC Chair and Trump Jr.'s Spokesman*, VICE (July 25, 2018), https://www.vice.com/en/article/43paqq/twitter-is-shadow-banning-prominent-republicans-like-the-rnc-chair-and-trump-jrs-spokesman.

[32] *See* Vijaya Gadde & Kayvon Beykpour, *Setting the Record Straight on Shadow Banning*, TWITTER BLOG (July 26, 2018), https://blog.twitter.com/official/en_us/topics/company/2018/Setting-the-record-straight-on-shadow-banning.html; *see also* Del Harvey & David Gasca, *Serving Healthy Conversation*, TWITTER BLOG (May 15, 2018), https://blog.twitter.com/en_us/topics/product/2018/Serving_Healthy_Conversation.

[33] *See* Brett Samuels, *Trump: 'We Will Look into' Twitter for 'Shadow Banning' Republicans*, THE HILL (July 26, 2018), https://thehill.com/policy/technology/398943-trump-government-will-look-into-twitter-for-shadow-banning-republicans.

[34] *See* Liam Stack, *What Is a 'Shadow Ban,' and Is Twitter Doing It to Republican Accounts?*, N.Y. TIMES (July 26, 2018), https://www.nytimes.com/2018/07/26/us/politics/twitter-shadowbanning.html.

[35] Blunt et al., *supra* note 10.

[36] *See* Megan McCluskey, *These Creators Say They're Still Being Suppressed for Posting Black Lives Matter Content on TikTok*, TIME (July 22, 2020), https://time.com/5863350/tiktok-black-creators.

[37] *Supra* note 16.

Today, the concept and terminology of "shadowbanning" are still central to debates about online content moderation. Only hours after buying Twitter, Elon Musk claimed that he would "dig into" into shadowbanning;[38] days later, Rep. Alexandria Ocasio-Cortez accused Twitter of shadowbanning her account for criticizing Musk;[39] and days after that, Musk tweeted out a new moderation policy that many argued amounted to shadowbanning, and then shortly deleted it.[40] Perhaps most significantly though, shadowbanning has recently entered the realm of legal contention. Shadowbanning is explicitly defined and prohibited in Florida's social media censorship law, which as of this writing is up for argument in the Supreme Court in *Moody v. NetChoice, LLC*. The Court will determine whether the First Amendment prohibits a state from requiring companies to notify users when they block their speech. Despite a dramatic change in the scale, design, and, perhaps soon, legal role of social media, the language we use to talk about undisclosed moderation is stuck on "shadowbanning.

## 2.    Why Retire "Shadowbanning"?

One major failing of the word "shadowbanning" is that its definitional slipperiness allows companies to outright deny that they shadowban yet continue to reduce the visibility of users' content without telling them. Instagram and Twitter, for example, have both denied that they shadowban. In an Instagram story, Instagram CEO Adam Mosseri said, "Shadowbanning is not a thing. If someone follows you on Instagram, your photos and videos can show up in their feed."[41] In a blog post, Twitter wrote, "People are asking us if we shadow ban. We do not."[42] Kelley Cotter calls this practice "black box gaslighting," and criticizes

---

[38] Sam Tabahriti & Sophia Ankel, *Elon Musk Says He'll Be 'Digging' into Shadowbans on His First Day at Twitter as Conservatives Urge Him to Overturn Their Previous Punishments*, INSIDER (Oct. 28, 2022), https://www.businessinsider.com/elon-musk-will-be-digging-into-shadowbans-twitter-first-day-2022-10.

[39] Graeme Massie, *AOC Accuses Elon Musk of Sabotaging Her Twitter Account: 'Doesn't Seem Very Free Speechy to Me,'* YAHOO! FINANCE (Nov. 3, 2022), https://uk.finance.yahoo.com/news/aoc-accuses-elon-musk-sabotaging-164501724.html.

[40] Nikki McCann Ramirez, *Right Wingers Fume After Musk Announces Shadowbanning Policy*, ROLLING STONE (Nov. 18, 2022), https://www.rollingstone.com/politics/politics-news/elon-musk-shadowbanning-policy-causes-right-wingers-fume-1234633498.

[41] Jesselyn Cook, *Instagram's CEO Says Shadow Banning 'Is Not A Thing.' That's Not True.*, HuffPost (Feb. 25, 2020, 1:41 PM), https://www.huffpost.com/entry/instagram-shadow-banning-is-real_n_5e555175c5b63b9c9ce434b0.

[42] *See* Gadde & Beykpour, *supra* note 32.

social media companies for using their "epistemic authority on their algorithms to undermine users' confidence in what they know about algorithms and destabilize credible criticism."[43] Whether explicitly or implicitly, in good faith or in bad, social media companies seem to be adopting the old, Reddit-era definition of shadowbanning rather than the more modern, inclusive one (any form of undisclosed content moderation), which they almost certainly practice.[44] Instead of divulging their rules for how they decide whether or not to inform users of a moderation action, platforms often categorically deny "shadowbanning" and move on.

In specific cases, undisclosed content moderation may be warranted. Social media companies need to be able to combat users that try to structurally abuse their services: disinformation botnets trying to circumvent anti-spam technology, trolls trying to learn what slurs trigger automated moderation, harassers creating new accounts to reach victims who have blocked them, and so on.[45] It is not clear, though, that social media companies limit undisclosed content moderation to just these cases. Even if they do, it would not justify hiding their policies about when they disclose moderation. Social media companies instead may be engaging in "constructive ambiguity."[46] Different stakeholders have different, sometimes mutually exclusive, goals for content moderation—for example, a government may want to crack down on dissenting speech while advocacy groups may want their dissenting speech to be heard. By providing incomplete information about their practices and policies, social media companies, if trusted, can convince both groups that their needs are being met.

Another problem with "shadowbanning" is that because it is both unprovable and undisprovable, it is less a subject of fact-based discourse

---

[43] Kelley Cotter, *"Shadowbanning Is Not a Thing": Black Box Gaslighting and the Power to Independently Know and Credibly Critique Algorithms*, INFO., COMMC'N & SOC'Y 1 (Oct. 28, 2021), https://doi.org/10.1080/1369118X.2021.1994624.

[44] *See* Laura Savolainen, *The Shadow Banning Controversy: Perceived Governance and Algorithmic Folklore*, 44 MEDIA, CULTURE & SOC'Y 1091 (2022); Erwan Le Merrer et al., *Setting the Record Straighter on Shadow Banning*, IEEE CONF. ON COMPUT. COMMC'NS (2021), https://arxiv.org/abs/2012.05101; Adam Mosseri, *Shedding More Light on How Instagram Works*, INSTAGRAM BLOG (Jun. 8, 2021), https://about.instagram.com/blog/announcements/shedding-more-light-on-how-instagram-works.

[45] *E.g.*, Kate Starbird, *Disinformation's Spread: Bots, Trolls and All of Us*, 571 NATURE 449 (2019); Luis Vargas, Patrick Emami & Patrick Traynor, *On the Detection of Disinformation Campaign Activity with Network Analysis*, ARXIV (May 27, 2020), http://arxiv.org/abs/2005.13466.

[46] Diane Coyle & Adrian Weller, *Explaining Machine Learning Reveals Policy Challenges*, 368 SCI. 1433 (2020).

and more a political cudgel levied against platforms. Among conservatives in particular, the rhetoric of shadowbanning has been used to make baseless claims of anti-conservative bias on Big Tech platforms.[47] Republicans in state legislatures have proposed over two dozen bills that specifically target shadowbanning,[48] including Florida's Stop Social Media Censorship Act, which passed in May 2021[49] and will soon be up for debate in the Supreme Court. Conservatives are not alone in feeling disproportionately targeted by shadowbanning. According to a survey from the Center for Democracy & Technology, 74% of social media users believe that shadowbanning unfairly targets certain groups more than others.[50] And users who believe that they have been shadowbanned are more often male, Hispanic, Republican, or non-cis-gendered.[51]

Even if the term was not so vague or politicized, platforms could never admit to "shadowbanning" because the term's connotations are simply too negative. The word implicitly casts social media companies as a sort of shadowy cabal, which may accurately reflect how users feel, but does not frame the issue in a way that encourages productive discussion of how companies *should* moderate content. As linguist George Lakoff might argue, the conspiratorial metaphor embedded in the word "shadowbanning" could reinforce actual conspiratorial sentiment, something companies certainly want to avoid.[52]

All of these shortcomings show that "shadowbanning" as a term has outlived its usefulness, and that it is time to find a more precise, less appropriable alternative. I personally recommend "undisclosed content moderation," a phrase I have used throughout this essay, which neatly encompasses both removal and reduction of content without casting platforms as occult villains. New language will enable more coolheaded

---

[47] *See* Paul M. Barret and J. Grant Sims, *False Accusation: The Unfounded Claim that Social Media Companies Censor Conservatives*, NYU STERN CTR. FOR BUS. & HUM. RTS. (Feb. 2021), https://bhr.stern.nyu.edu/bias-report-release-page.

[48] *See Search Results for "Shadow Ban,"* OPEN STATES, https://openstates.org/search/?query=%22shadow+ban%22 (last visited Oct. 29, 2022).

[49] *See* Sofia Andrade, *Florida's New Pro-Disney, Anti-Facebook and Twitter Law*, SLATE (May 25, 2021), https://slate.com/technology/2021/05/florida-stop-social-media-censorship-act-disney.html.

[50] *See* Nicholas, *supra* note 16.

[51] *See id.*

[52] GEORGE LAKOFF, DON'T THINK OF AN ELEPHANT! KNOW YOUR VALUES AND FRAME THE DEBATE (2004). Not all linguists agree with the importance Lakoff puts metaphors. For criticism of his "generative semantics" theory, see Steven Pinker, *Block That Metaphor!*, THE NEW REPUBLIC (Oct. 9, 2006), https://newrepublic.com/article/77730/block-metaphor-steven-pinker-whose-freedom-george-lakoff.

public debate about when not telling users that their content has been moderated is justified (*e.g.*, against inauthentic coordinated actors) or not (*e.g.*, against explicit content). But shifting to this decidedly blander nomenclature is not up to end users like Jessie Sage, for whom "shadowbanning" is currently the only commonly understood language available to discuss their experiences. Instead, it is primarily up to social media companies, who advocate for new vocabulary by using it to fully disclose their rules for when and how they inform users that that they are removing or reducing the visibility of their content. By describing these conditions, platforms can shift what has been called "shadowbanning" from an unknown unknown to a known unknown, making it less amenable to evasive public relations maneuvers and political weaponization. Platforms certainly have the rhetorical might to push this kind of change. They have exercised it before popularizing new buzzwords, like "metaverse," and even new language pertaining to content moderation, such as "harmful content," "toxicity," and "borderline content.

   "Shadowbanning" is a term and a practice born from the nascent days of the internet. Platforms were small and unruly, and early site administrators built themselves powerful, opaque tools to keep their communities in check. But "shadowbanning" stuck around long after those halcyon days ended. As social media grew into arguably the predominant means of human communication, platforms' decisions about when and how to inform users of content moderation took on democracy-sized consequences. Today, the language of "shadowbanning," perhaps once playful, is no longer up to the task of upholding public discourse about what those practices should be, and this failure hurts users and platforms alike. If social media companies want a word to describe undisclosed content moderation that doesn't cast themselves as shadowy puppeteers, they will have to use it to describe their own moderation practices. In other words, to end the word "shadowbanning," social media companies may need to be just a little less shadowy.