

## Identifying Information

Name:	Goldman, Eric
School:	Santa Clara Law

## Paper Information

Title:	Content Moderation Remedies
Abstract:	<p>How online services moderate content—i.e., how online services decide what third party content to publish or remove—has become a major social issue, prompting substantial news coverage, Congressional hearings, and proposed regulatory interventions. Most of the regulatory discussion focuses on the substantive content rules, i.e., what content is fit to be published. There has also been some discussion about how online services should operationalize those content rules. This article takes a novel approach to supplement the existing discussion by focusing on the remedy question. Online services have the technical and legal power to impose remedies for rule violations, and they can impose these remedies extrajudicially by effectuating them within their network. Because these remedies function outside the formal legal system, the online services have substantial flexibility in designing and implementing their remedies for problematic content and users. So, what remedies should they impose on content or user that violates their substantive rules? And what design principles should they adopt when developing and implementing those remedies? Most Internet regulations don't address the remedies for content moderation at all. In the rare circumstances where the regulators address the topic, they treat remedies as binary: content stays up or gets taken down; user accounts remains active or are terminated. For example, the DMCA online safe harbors for third-party caused copyright infringement illustrates the regulators' binary conceptualization of remedies. 17 USC 512(c) conditions the safe harbor on "removing" or disabling access to" the allegedly infringing files and "terminating repeat infringers." In fact, online services have a long list of potential remedies available to them beyond the binary extremes. I've identified about two dozen different remedies that online services can impose on problematic content or accounts. These remedies can be organized into five major categories: (1) reduced content visibility, (2) content restrictions, (3) demonetization, (4) account restrictions, and (5) others. Not all remedies fit all circumstances or services; for example, demonetization options don't work if the online service doesn't pay users for content. Nevertheless, most of the remedies I've identified have been deployed by at least one online service and could be adopted more broadly. By taxonomizing and describing the remedy options, the article will help regulators recognize the richness of the issues and the wider range of options they can pursue; and the list should spur brainstorming among industry participants. The article will also address, for the first time, how online services should design and structure their remedies for problematic content and users. Assuming the law gives them flexibility, the article will advocate that online services should prioritize the enhancement of user speech and design a remedy set that supports this goal. For example, rather than deleting content, the online services can opt to reduce the visibility of problematic content whenever possible (and where consistent with the legal requirements). Diminishing content visibility can promote user speech by enabling niche or minority communities to communicate with each other without disturbing the service's main audience. By identifying the availability of a wider range of remedies for problematic content or speech, the article will help the online services—and regulators—develop more nuanced and fine-grained outcomes than are available with the</p>

	binary extremes. This, in turn, will result in more user speech avoiding online “death penalties” while still supporting the service’s other content moderation goals.
--	--