The Virtues of Moderation: Online Communities as Semicommons

James Grimmelmann*

IP Scholars workshop draft: not intended for broad distribution or citation

* Associate Professor, New York Law School. My thanks for their comments to Jack Balkin, Shyam Balganesh, Aislinn Black, Anne Huang, Amy Kapczynski, David Krinsky, Chris Riley, Steven Wu, and the participants in the May 2007 Commons Theory Workshop for Young Scholars at the Max Planck Institute for the Study of Collective Goods.

I. Introduction

What lessons should intellectual property law learn from the flourishing of online communities? One answer, associated most prominently with Yochai Benkler, is that Internet-facilitated collaborative communities can perform the creative work intellectual property laws seek to encourage, but without needing to rely on the exclusionary controls that characterize intellectual property regimes. This view sees online communities as successful commons. Another, more skeptical view, is that a "successful online commons" a chimera; any online community is either not a commons or will not be successful for long. Notice the shift. The policy question—*when and where is intellectual property appropriate in the Internet age?*—has become, in part, a descriptive question—*what is really happening in online communities?* And that question, in turn, is really a question about what we mean when we say "commons" and "online" in the same breath. In this paper, I will argue that when we do, we should really be saying "semicommons."

The standard economic explanation for imposing controls on people's use of resources is that doing so discourages waste and encourages investment. Private property is the classic example of such control; direct governmental regulation was traditionally seen as its alternative. Many authors have assumed that one or the other would be necessary to prevent disastrous misuse. But scholars of the commons have told two stories that question this traditional assumption.

One the one hand, the rich literature on common-pool resources has demonstrated that property and government are not the only possible sources of control. In addition to property and regulation, both of which are enforced through top-down legal commands, it is also possible for communities to monitor and police their own usage of a resource through bottom-up, self-enforced controls. This story focuses on the problem of waste, considers some form of control to be a useful way to solve that problem, and explains that common ownership is most effective when the relevant community is small, well-defined, and close-knit. Call this story the Tragic one; it is essentially pessimistic.

On the other hand, the growing literature on the intellectual commons has built a powerful case that extraction of economic rents is not the only motivation to contribute to a resource; sociability, self-improvement, signaling, and gains from related resources are also

important incentives. Thus, widespread freedom of use, in part by forgoing the transaction costs associated with a system of control, can harness these other motivations to achieve large-scale collaboration. This story focuses on the problem of investment, considers forgoing control to be a useful way to solve that problem, and explains that common ownership is most effective when the relevant community is large, diffuse, and diverse. Call this story the Comedic one; it essentially optimistic.

This disconnect between these two views of the commons is at the heart of recent economic debates over online communities. Comedic authors, noting the remarkable success of open source software, of Wikipedia, and of other communities built around freely-shared information, have argued that we are entering a new age of collaboration and cooperation. Since incentives to contribute will only grow with more contributors, the sky is the limit. Tragic skeptics, on the other hand, warn that past performance is no guarantee of future results. The spectacular cheapness of computer technology has only deferred congestion problems, not solved them. As unsophisticated users, malicious competitors, and greedy spammers rush in, familiar problems of waste will reemerge. There would seem to be a fundamental conflict between the Comedic claim that social incentives make freedom irresistible and the Tragic claim that the specter of waste makes control inescapable.

Both are right; they have focused on different aspects of the problem. An online commons is really two inextricably conjoined resources with very different properties: information, and the substrate used to communicate it. The Comedic view correctly argues that because the information is nonrival, there are good reasons to want to make access to it as free and open as possible; doing so strongly encourages its production. The Tragic view correctly argues that because information is communicated using rival resources, there are also good reasons to restrict access to these underlying resources; doing so discourages their wasteful overuse. The challenge is to balance these two imperatives. Thriving communities hew to a middle ground in which much is permitted and some is not. Their experience shows that partial and well-chosen controls can thwart gross misuse while preserving many of the rich set of freedoms that make for a healthy community. The result is a semicommons, in which the conjoined resource set is partly private and partly held in common.

Taking this view of online communities has several useful consequences. First, by characterizing them both as private and as commons, rather than one or the other, it focuses

attention on the forms of moderation that communities use to guide discussions towards fruitful outcomes, direct attention where it is needed, prevent disruption, and coordinate collaboration. This diversity of institutional responses is itself significant; online communities cannot be reduced to a single axis from openness to closure. Second, and similarly, it emphasizes that the success of collaborative communities is context-dependent, so that subtle differences can make the difference between fruitful order and disruptive chaos. Third, it makes visible the dynamism of many successful online communities in repeatedly devising moderation strategies to deal with changing circumstances. Successfully maintaining the interplay between private and common requires not one precisely-adapted set of patterns, but rather the ability to adaptively forge new ones. This last point provides a new perspective on the generativity of especially successful online semicommons communities, such as Wikipedia and the Internet itself.

Part II of this paper will review past work on commons theory and explain in more detail the concepts behind the Tragic and Comedic stories. Part III will combine these stories in an abstract model of an *online semicommons*, and Part IV will use that model to extend past work on online communities and resources. Part V will present three case studies of moderation in action.

II. Background

A. Public and Private Goods

Economic analysis of the commons conventionally begins with the distinction between *private goods* and *pure public goods.* Private goods such as cars, farms, and handbags have two key characteristics. First, they are *rival* (or, sometimes, "subtractable"): one person's use of the good makes it unavailable for someone else. Only one person at a time can drive a car; only one person at a time can wear a handbag. Second, they are *excludable*: it is possible to prevent people from using the good. We can put locks on cars and fences around farms. These characteristics have two salutary effects. First, they discourage wasteful uses. If I value a good more than you do, it will be profitable for both of us for me to buy it from you. Excludability makes this exchange meaningful and effective; otherwise, I could simply take the good from you and you could take it back, ad nauseam. Second, they promote efficient investment in creating new goods. Provided some simplifying assumptions hold, rivalry and excludability permit me to arrange the potential buyers' individual valuations into a demand schedule; I will then produce

the good until the marginal buyer's willingness to pay exactly equals my marginal cost of production.

Interesting and important problems arise when either rivaly or excludability is absent. Goods that are rival but not excludable are *common goods*. These goods are subject to wasteful racing behavior as different potential users seek to make use of them before someone else does. The result is the so-called "tragedy of the commons," as the resource is depleted through competitive overuse. This depletion is fundamentally a problem of nonexcludability; where competing users can be kept from the good, the race can be terminated by whoever has the power to exclude. The classic example of such a common-pool resource is a communal pasture on which many different shepherds graze their sheep. The field's supply of grass is limited, but no shepherd can keep his neighbors from adding more sheep to their herds. The rational shepherd is led to add more and more sheep to grab a larger share of the pasturage, even as he expects the others to do exactly the same. The traditional response is to find some way of regenerating excludability. Private property rights are one such way, but many common pool resources are structurally resistant to such rights. It may be hard to define the scope of a property right in fish in the sea, or to enforce a property right against taking oil from an underground pool. Government regulation is another solution, but the government may be no better at direct monitoring than it would be at enforcing individual property rights. As we shall see, bottom-up community management can sometimes go where these other institutions cannot in generating a manageable regime of exclusion.

On the other hand, goods that are nonrival generate a problem of undersupply. On the standard analysis, the provider of a nonrival good has positive fixed costs to produce the first copy of the good, but zero marginal cost to produce subsequent copies. The normally efficient strategy of setting the price equal to her marginal cost is efficient only once she has decided to produce the good in the first place. If the good were excludable (a *toll good* or *club good*) and she could perfectly price discriminate, then she would have efficient incentives to produce exactly those goods valued by her buyers. In practice, however, perfect price discrimination is usually impossible, and buyers have strategic incentives to understate the value they place on the good, in hopes that others will pay enough to cause it to be produced.

This free-rider problem is also present when then good is both nonrival and nonexcludable (and is thus a *pure public good*), but with a different twist. Without excludability,

there is little possibility of a market in such goods; as soon as more than one person has the good, competition between them will drive the price to zero. Traditionally, there have been two interesting responses. The first is to invest in making the good more excludable, in the hopes of creating a toll good for which the production incentives are at least present, even if inefficiently sized. This is one explanation of copyright, for example, which creates a form of legal excludability. The exclusion is typically costly, and so the costs of exclusion must be balanced against the incentive gains for production. The other response is to embrace nonexcludability and for the government to invest directly in funding the resource. This response has been used to support government production of lighthouses and defense, as well as prizes for innovation. This approach can take advantage of the economies of scale enabled by nonexcludability, but faces the difficult problem of finding the optimal level of funding for this investment. Once again, the key issue is to learn the valuations individuals would place on the good.

To sum up, resources with one or both characteristics of pure public goods—nonrivalry and nonexcludability—can present two distinctive problems. First, there is the problem of waste, which is most acute when the good is rival but nonexcludable. Second, there is the problem of insufficient incentives to invest, which arises when the good is nonrival and was traditionally considered most acute when the good was also nonexcludable. Thus, prior to the modern flourishing of commons theory, one might naturally have believed that the best response to both problems was to focus on creating excludability. In one context, commons theory has embraced that response; in another, commons theory has repudiated it. It is to these two stories that we now turn.

B. The Tragic Story

In 1968, Garrett Hardin's influential article "The Tragedy of the Commons" appeared in *Science* with essentially the argument given above about wasteful overuse of common-pool resources.[1] His specific subject was overpopulation, but he presented examples from parking meters to pollution and delivered the sobering verdict that:

> Adding together the component partial utilities, the rational herdsman concludes that the only sensible course for him to pursue is to add another animal to his herd. And another; and another. . . . But this is the conclusion reached by each and every rational herdsman sharing a commons. Therein is the tragedy. Each man is locked into a system that compels him to increase his herd without

[1] Garrett Hardin, *The Tragedy of the Commons*, 162 SCIENCE 1243 (1968).

> limit--in a world that is limited. Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the commons. Freedom in a commons brings ruin to all.[2]

Against this tragic fate, Hardin saw only one effective response: coercion. "[T]he necessity of abandoning the commons"[3] would require "mutual coercion, mutually agreed upon"[4] to establish effective restrictions on overuse. Such restrictions might take one of two forms: "private property" in the formerly common resource or "allocat[ion] of the right to enter."[5] On this view, continued common ownership of a resource would lead inexorably to the tragedy of exhaustion.

Subsequent research on common-pool resources has challenged this last conclusion. A rich literature has found that many natural and man-made resources are in fact held communally (whether by law or in practice) and have in many cases been held in common for many hundreds of years without suffering this supposedly inexorably tragic fate. The modern synthesis view is that such cases involve communities that have succeeded in creating and then enforcing on themselves a regime of limited access to the resource. Unlike property rights or direct regulation, both of which are institutions supplied by the state, these community management regimes are bottom-up and are at least initially the product of voluntary cooperation. Nonetheless, they do effectively supply Hardin's "coercion;" community members who violate the community's access rules will be punished by they community. I have called this revised story of common-pool community management the Tragic story because it remains within the basic parameters of Hardin's explanation of the Tragedy of the Commons. It concedes that a common-pool resource will be tragically overused unless some source of control over its use can be found; it claims, however, that common ownership and the necessary control can be compatible.

This revised Tragic story matters for several reasons. First, it is a caution against some ill-advised forms of state intervention. The common users of a resource may be more attuned to its actual characteristics, more able to monitor its usage, and more able to apply appropriate sanctions than a governmental actor would be. Policy-makers who know their Hardin might conclude that they are compelled to privatize or micromanage a common-pool resource; those

---

[2] *Id.* at 1244.
[3] *Id.* at 1248.
[4] *Id.* at 1247.
[5] *Id.* at 1245.

who know their Ostrom as well would recognize that doing so could well destroy the local knowledge that actually makes successfully restrained usage possible. Second, there can be important reasons over and above avoiding tragedy to desire common ownership. Common ownership can have very different distributional consequences than private or government ownership; it can create substantial risk-sharing within a community and can smooth some inequalities. Similarly, the bottom-up process of creating and enforcing usage policies serves important participatory values and can do so more directly than parcipation in government policy-setting would. These distributional and participatory values are among the reasons for the resonance of the phrase "enclosure of the commons"; the transformation of a common resource to a privately-held one can both immiserate and disenfranchise many community members. Finally, and most importantly for present purposes, the Tragic story indicates the vitality of commons-based community self-governance systems of resource management. They are neither archaic holdovers nor illusory bulwarks; under the right circumstances, common ownership can be an effective mechanism for adapting resource usage to a community's short-term and long-term needs.

Not all circumstances are right. The literature on common-pool resources also is full of examples of attempted common ownership regimes that have failed. Exact lists differ, but there is by now a rough consensus on the core factors that distinguish successful from unsuccessful common ownership regimes. The community needs both good institutions to gather information about the resource and its usage and a forum to discuss its management. There should be graduated sanctions available, starting with small shaming penalties for minor or first infractions against the access rules and escalating to severe penalties for sustained and flagrant overuse. The community must be able to participate in making and enforcing the rules, at multiple levels of generality. And, perhaps most critically, the community itself should be well-bounded, so that the set of potential users is known, mostly closed to new entrants, and ideally small.[6] Ostrom considers this last condition essential. It enables a group to make the substantial investment in a management regime without fear that outsiders will come racing in and undo all their work. While large common-ownership groups do sometimes exist, they typically involve the

[6] Homogeneity of membership, once thought to be critical as well, no longer has this pride of place, and many scholars now think that even substantially heterogeneous groups can achieve the necessary cooperation if other mechanisms of mutual assurance are properly in place.

aggregation of smaller well-functioning units. The small, close-knit group is the model for common-pool resource theory; large and diffuse groups are generally seen as poor candidates for common ownership. The expectation is that they will fail to solve collective action problems and will fall back into wasteful overuse. Remarkably enough, another strain of commons theory has come to almost exactly the opposite conclusion.

C. The Comedic Story

I have discussed briefly above the conventional economic apology for intellectual property, on which the nonrivalry of its consumption creates difficult problems in setting the proper level of incentives for investment in its creation; legally-enforced exclusivity is a device to move those incentives closer to the optimal level. This view has always coexisted with a strong critique: once an (idealized) intellectual good exists, its nonrivalry means that it can be shared with the world at no cost, and its natural state of nonexclusivity means that it will be. Thus, the legal exclusivity at the heart of intellectual property law is an artificial state-created scarcity. Further, from this *ex post* perspective, no further incentive is needed for the creation of the good once it exists. Thus, seen in hindsight, the artificial scarcity of intellectual property is an inefficient wealth transfer to creators and at times an offense against distributive justice. The conventional narrative of intellectual property law synthesizes these two opposite arguments into a dialectic that pits *ex ante* incentives to create against the *ex post* value of broad access. Policy-makers are told to set the level of legal protection to maximize the overall distribution of works to the public; the level of protection should be neither too high nor too low.

There is, however, a deeper critique. Intellectual goods are not simply consumed. They are also critical inputs into the production of other intellectual goods. All creativity is deeply influenced both by specific inspirational earlier works and by the entire cultural environment within which a creator works. All innovation is incremental and builds on past inventions while looking for solutions to the new problems that past discoveries open up. On this account, overprotection poses a critical additional danger to future creativity, and lower protection can increase creativity by speeding the circulation of ideas. The public domain becomes not simply a negative space of unprotected works and inventions, but a positive resource of immense richness that is and should be held open to all: a commons. Indeed, because there is no problem of rival consumption and thus no danger of waste, it becomes a perfectly functioning commons, instead of a depletable common-pool resource susceptible to exhaustion.

This account has borrowed from the commons theory literature on tangible resources in two important ways. First, it conceptualizes the problem created by overprotection as an *anticommons*, a space in which many actors have the ability to prevent action, so that nothing is ever done because the transaction costs of negotiating all of the necessary permissions would be prohibitive. If we were required to obtain permission to use each word we speak from the heirs of the first person to use it, language itself would collapse. This tragedy of the anticommons is the reciprocal of the tragedy of the commons; it provides a powerful economic and rhetorical counterargument to excessive—or at least inappropriately calibrated—propertization of a commonly-held resource. Underuse can be just as wasteful as overuse, and strategic holdout behavior is a constant threat when any of a number of owners can veto a proposed use.

Carol Rose's remarkable *The Comedy of the Commons*[7] makes a version of this point about holdouts and also offers a critical observation about social uses of resources. By looking at the historical law of "inherently public property," particularly roads and waterways, Rose developed a theory of the role of the public (distinct from government) as an owner and manager of property. She pointed to the scale returns (also called "network effects" or "positive network externalities") associated with increased use of certain public spaces. A dance in a public square becomes more enjoyable for all participants as it becomes more popular; commerce along a road network becomes more fruitful for all traders as the number of trading partners increases. She then combined this observation with the holdout point to create a positive theory of the benefits of common ownership:

> Suppose that a private individual owned a traditional festival ground and that, at least for the festival day, the local residents placed a higher value on this festival use than could be taken from any alternative uses. Ownership of this unique property would give the owner a classic opportunity for rent capture.
>
> But what created the "rent"? The very "publicness" of the festival use; its non-exclusivity makes it valuable, because this activity is exponentially enhanced by greater participation. This value is what customary doctrines refused to permit a private owner to tap or to thwart. In fact, the usual rationing function of pricing would be counterproductive here: participants need encouragement to join these activities, where their participation produces beneficial "externalities" for other participants."[8]

---

[7] Carol Rose, *The Comedy of the Commons: Custom, Commerce, and Inherently Public Property*, 53 U. CHI. L. REV. 711 (1986).
[8] *Id.* at 768–69.

This is an important inversion. It is not just a theory about the dangers of holdouts; it is also a theory of *self-generated incentives for investment*. If the holdout problem can be solved and the costs of participation driven down, the value created for each individual by contact with others becomes sufficient incentive for her to participate. The added value to others created by her participation invites further participation from them, and so on in a virtuous cycle.

This point has resonated powerfully with theorists of the intellectual commons. Writers such as Yochai Benkler, Rishab Ghosh, and Steve Weber have applied this mode of analysis to open source software and other "impossible public goods." They have argued that voluntary contribution to an intellectual commons can indeed be self-sustaining and that Rose-style scale returns are one important factor in overcoming the seemingly intractable incentive problem. They have also added factors of their own. Many such projects are organized to split tasks up into small and easily manageable components. Even without the ability to charge for their contributions, participants can realize gains from complementary resources, from signaling their interest and skills, from the pleasure of creation and participation, and from the value of the product itself to them. These incentives may not necessarily be large for many participants, but they are large enough to elicit participation, which is all that matters. In all of this, common access is critical. Common access cuts with the grain of the intellectual good's nonrival and nonexclusive character, rather than against it. Common access lowers costs beneath the threshold at which broad participation becomes feasible, and thereby harnesses scale returns from increased participation. Enforced common access also contributes to shared group norms of a common productive (and altruistic) enterprise and provides a partial guarantee that contribution will be reciprocated, rather than appropriated for the sole benefit of one owner.

Thus, to recap, the intellectual property strain of commons theory now has a narrative about free and open access; in Rose's phrase, "the more the merrier." Since the goods involved are nonrival, free riding poses no threat of waste and broad access serves goals of efficiency distributive justice. A large intellectual commons, particularly in the form of the public domain, promotes creativity by many, which feeds on itself, and leads to broad and democratic participation in culture and innovation. These benefits are squandered if access is restricted or a gatekeeper is empowered to exclude others from the intellectual commons. It is common access that makes the machine go of itself, and the machine goes better the more widely and cheaply the access is opened. Commons, (or "social" or "peer") production will coexist with market

production, rather than displacing it, but in many domains there are enormous possibilities open to do things better than they have been done thus far. All that is required is to throw the gates open, to make the common community of creators and users as large as possible.

D. Layering

Thus, one strain of theory says to restrict access to a small community; another says to make the community enormous. There is no direct conflict: the first strain by its terms applies to rival resources and the second to nonrival ones. This is all well and good, but the most pressing problems for those who care about such things do not involve resources that can be easily placed into one bin or the other. Indeed, the pure theory of the intellectual commons only directly applies as such to information itself. Given that information must always be instantiated to be used or communicated—whether noted in a human mind with limited attention, stored on a tangible object, or transmitted through an online channel with limited bandwidth—the Tragic story is always close at hand, ready to apply to these rival instantiations. At the same time, it seems clear that something more is going on than simple competition to exhaust the rival components of instantiated information. Scholars have developed one further idea that preserves the possibility that both stories could apply simultaneously: layering.

The term comes from computer science, where it applies to the different "layers" involved in a computer network. Two computers may use SMTP to exchange email over a network that runs TCP/IP over 1000BASE-T gigabit Ethernet over Category 5e twisted pair copper cables, but these physical and lower-level networking details are irrelevant from the point of view of the email program that the owners of these computers use to exchange messages. The wires (the "physical" layer), the Ethernet protocol (the "link" layer), the TCP/IP protocol (the "transport" layer), the SMTP email protocol (the "application" layer), and the contents of the emails themselves (the "content" layer) are distinct. Ethernet uses a clever system, known as exponential backoff, to ensure that only one computer is sending a message at a time, but the email program can work perfectly well despite having no knowledge of how exponential backoff works—or even that the network, multiple layers down, uses Ethernet. Layering is a form of modularity—breaking a system down into subcomponents that do not need to know the details of how the others work—and it offers many of the purely technical benefits of modularity.

More than that, however, layering also also permits different resource allocation regimes at different layers. In particular, one can make a system controlled at one layer but free at a

higher layer. The same network may be fully private at the physical layer (only the company IT manager can enter the room with the server), a limited-access common-pool-resource at the link and transport layers (only employees in the building can connect to it, but they can do so freely), an open-to-the-world common-pool-resource at the application layer (the company provides free ad-supported image hosting to Internet users at large), and something approaching a true commons at the content layer (the company exercises no control over what images users create and share). This form of layered sharing is visible in many communications systems, but is a particularly prevalent feature of the Internet. Indeed, a number of scholars give it credit for the Internet's remarkable success. They see an application layer largely open to new innovations and a content layer largely open to all forms of communication and sharing, both shining examples of the Comedic story.

Of course, one cannot manufacture something from nothing and exceed one layer's capacity simply by adding more layers atop it. Where capacity is a serious problem, the Tragic story is also plausible. Simplifying greatly, one might say that many debates over Internet law and policy have been driven by a debate between those who tell a Comedic story about the higher layers and those who tell a Tragic story about the lower ones. Comedians consider the productive freedoms at the higher layers so valuable that they require legal protections, with waste problems at the lower layers either practically irrelevant due to surplus capacity or a proper subject for specific, targeted responses that leave higher-level freedoms intact. Tragedians instead see the lower-level waste issues as fundamental, so that providers will be unwilling to invest in creating capacity unless they have sufficient legal protection. This dynamic is clearly visible in the debates over spectrum allocation, network neutrality, and trespass to computer systems. Comedians fear that infrastructure will try to capture the gains from the higher-level commons, thereby killing the goose that lays the golden eggs; Tragedians fear that if there is too much solicitousness of the higher-level commons there will be no infrastructure.[9]

I do not intend to enter into these debates as such. Instead, I would like to focus more closely on the interactions between layers—in particular, on those mechanisms that moderate use of the higher layers. Almost every communications system, even the most seemingly unrestrained, has some such mechanisms in place, ranging from highly informal social norms to

[9] This debate is visible in Rose.

quite detailed and mechanistic pricing schemes. These forms of moderation are best understood as techniques to prevent Tragic overuse of the higher layers in ways that would tax the capacity of lower layers. At the same time, in many of the most successful online communities and communications systems, the moderation is remarkably lightweight; the Comedic virtues are slightly compromised, not abandoned.

III. Online Semicommons

The three key pieces of theory presented in the previous Part apply to many familiar online communities: they combine a potentially rival communications layer with a generally nonrival content layer. Wikipedia runs on real servers and runs up an enormous monthly bandwidth bill. Nonetheless, anyone with an Internet connection can read and edit its entries. The same pattern is so frequently repeated elsewhere that we have become insensitive to just how remarkable it is. Mailing lists with thousands of subscribers are not overwhelmed with traffic. Juvenile jostling and nerdy one-upsmanship somehow combine to make Slashdot a must-read news site for technophiles. The average communication on Myspace is barely indistinguishable from the output of a Markov-chain random-phrase generator, and yet it has become immensely popular with its users. Spam has not driven blog comments, or blogs themselves, from the Internet. Indeed, people can still generally find what they're looking for on the Web, even though anyone can attach a server and start trying to game Google's rankings. Computers, bandwidth, and human attention are far from limitless and hardly free. And yet, time and time again, when these private goods are hooked together and common access to them granted at a higher layer, absurdly valuable public goods emerge. Something subtle is happening here, something that moderates use without overly suppressing the valuable properties of openness. This Part will set up the necessary analytic machinery to explain what that "something" is.

A. A Formal Model

We begin with a formal model. This model will apply to discussion lists hosted on a single server (such as the Cyberprofs mailing list), to complex user-generated-content sites hosted on an entire server farm (such as Wikipedia), and to distributed systems running common protocols (such as the entire SMTP-based email system). Henry Smith's concept of a semicommons provides the natural abstraction:

> In a semicommons, a resource is owned and used in common for one major purpose, but, with respect to some other major purpose, individual economic units—individuals, families, or firms—have property rights to separate pieces of the commons. Most property mixes elements of common and private ownership, but one or the other dominates. . . . In what I am calling a semicommons, both common and private uses are important and impact significantly on each other.[10]

Smith's "archetypal example of a semicommons" is the common-pasturage system of medieval Europe. Sheep could be grazed freely across the fields of a village during fallow seasons, but during growing seasons, individual farmers had exclusive rights to their strips of land. Smith's framework explains both the positive value created by enabling different access regimes for different uses and the need for governance rules to prevent strategic behavior by private and common users in placing costs on each other. This framework maps naturally onto the problems faced by many online communities, where the physical layer is privately owned but the content layer is effectively held in common.[11]

I call the result an *online semicommons*,[12] in which a community of users participate by exchanging content using infrastructure provided by one or more owners. The same person may of course be both a user and an owner. The content is nonrival and not naturally excludable; the infrastructure is both rival and excludable. Users participate in three ways: by writing, by reading what others have written, and by engaging in moderation. The same person may engage in all three activities, and there is no point in being overly rigid about the distinctions. Authorship and moderation are closely allied and there are systems in which the two are virtually indistinguishable at the border. Readership and authorship are also frequently two sides of the same coin; a reader will be inspired by another's contribution to create one of her own. This tight feedback cycle, in turn, can look much like moderation. The point is simply that the life cycle of a contribution involves its creation by an author, its alteration by moderators, and its

---

[10] Henry E. Smith, *Semicommon Property Rights and Scattering in the Open Fields*, 132 J. LEGAL STUD. 131, 132 (2000).

[11] The network layers in between are designed to give users access to the commonly-held content, but to deny them more fundamental powers, such as the power to take the server entirely offline; they therefore define the boundary between the common and private aspects of the system

[12] Robert Heverly has written of the "information semicommons," but he refers, instead, to the interplay between private and common uses of information. Copyrighted information is privately owned; the public domain is held in common. Robert Heverly, *The Information Semicommons*,18 BERK. TECH. L. J. 1127, 1164–72 (2003). Heverly is thinking about all human use of information My focus is narrower in that I look only at practices within a given online community, but broader in that I include the community's infrastructure as part of the relevant resource set.

consumption by readers. With that point in mind, I make the following assumptions about the costs and benefits of participation:

- Authors may have some intrinsic motivations to write that are independent of the size of their audience, but they primarily value being read. (More readers means more fans, more influence, and more altruistic satisfaction.) On the other hand, writing takes time and effort. The size of the audience does not directly affect an author's costs of writing, but the prospect of a larger audience may be a spur to greater effort.
- Moderators, like authors, care that particular items of interest to them be read by others. Moderators may disagree about which items are worth promoting. I will call a moderator's preferences among items her "ideology." [13] Moderation, like writing, takes time and effort. Some of that effort is spent in reading enough to decide what moderation is necessary; some of it is spent actually flipping the necessary switches.
- Readers gain value from reading particular items of interest. While writers are generally happy to have a larger audience and are comparatively indifferent about its identity, readers are highly idiosyncratic in their preferences across authors. Different readers may or may not value a given contribution identically. Reading an item in full takes time and effort. By default, readers have no way of deciding which items to choose to read, but moderation can help them pick and choose.
- All three activities impose costs on the infrastructure, which is a rival resource that is costly to provision and could be used for other purposes.[14] These costs are borne by the owners.[15] Writing is generally substantially more costly for owners than moderation or reading, usually because adding a contribution requires

---

[13] One could also say that authors have an ideological preference for their contributions to be widely heard. I do not think that much hangs on how exactly we characterize authors' preferences; the shape of the resulting curve is the same.

[14] On the nature of those costs, *see* infra Part IV.C.1.

[15] I am not including the limited time and attention of the community members in the privately-held portion of the communication semicommons. I model those effects in the participants' cost functions. I believe that these factors could be rolled into the description of the infrastructure without affecting anything essential in my argument; I have not done so because it seems unintuitive to talk about attention as a shared resource. Smith does not model the farmer's or shepherd's labor as part of the open-fields semicommons; I similarly think that participants' labor is not part of an online semicommons as such.

allocating some storage capacity for the indefinite future. The private owners can recover some of these costs by charging users, either explicitly (e.g. a monthly subscription fee) or implicitly (e.g. by showing advertising). They may also benefit in the same way that moderators and authors do: through goodwill, altruism, and publicity.

- Finally, because the same person could be an author, reader, moderator, and owner, these cost-and-benefit functions interrelate. An author might be described as having an ideological moderation preference for her own writings. All participants may all be motivated their own interest as readers: the community produces something valuable to them. (Thus, for example, in may peer-to-peer networks, users supply their own infrastructure so they can download the files they want. Similarly, an author might participate in an open-source project because the finished product will be valuable to her and she would like a say in how it comes out.) Private ownership brings with it control of the lower layers, and thus increased moderation power, which can drive users to become owners, and owners to become moderators.

B. Commons Virtues

Let us take a step back and ask what is truly characteristic about the commons aspect of this model. At the end of the day, users of any online community will have some range of freedoms and produce some set of information goods. We would like to be able to evaluate these freedoms and those outputs to ask how successful the community is as a commons. This section offers one such set of criteria: a list of commons virtues. These virtues are defining characteristics of an ideal commons. While no actual institution achieves even one of them perfectly, they can serve as yardsticks to measure how free a community is and how successful at harnessing collaboration.

First, access to an ideal commons is *broadly open*. The set of participants in the community should include as many as possible of those who would like to participate. No one should be excluded. A wiki usable by anyone on the Internet is more open than a wiki open to anyone on a school's network, which is more open than a password-protected wiki open only to the graduate students of the geology department. Specific failures of openness—in which particular individuals are singled out for exclusion, and especially for invidious reasons—may be

especially troubling. Openness applies to all three roles of participation; anyone should be able to write, anyone should be able to read, and anyone should be able to moderate.[16] A community may be open with respect to one but not another; for example, many moderated mailing lists are completely open to readers, moderately open to authors (whose posts go through a moderator but are usually approved), and open only to a single moderator.

Second, access to an ideal commons is *cheap* (i.e. close to zero cost). Openness and cheapness are independent. A fancy restaurant may be open to all, but its high prices will still exclude many. Many moderated social communities, on the other hand, are completely closed to outsiders but are cost-free to their members. Both exclude, but in different ways. Not all costs are explicit, and this definition of cheapness includes various imputed costs of participation. Thus, it includes the cost of the effort involved in writing, moderating, and reading. Like broad access, cheapness has a distributional dimension. A cost might fall on authors, readers, moderators, or infrastructure owners. Metafilter is free to read, but it costs $5 to sign up for an account that allows one to post.

Third, an ideal commons is *productive*: it generates valuable information goods. Authors find the commons productive when they can reach an audience; the larger the better. Readers find it productive when they can find high-quality information goods in it. Moderators find it productive when they can participate in the distribution of information goods they approve of. And society at large will find it productive when the spillover benefits of the information produced outweigh any negative consequences for others. (A community dedicated to trading stolen credit-card numbers would be unfortunate, even if its users would consider it productive.)

Finally, an ideal commons is *democratic*: its users participate in the community's self-governance. There are as many justifications for this virtue as there are theories of democracy. I will discuss below some of the instrumental reasons for users to take part in determining the community's choice among patterns of moderation. There are also substantive reasons to prefer self-governance. Participation is itself a source of satisfaction. It enables the individual to understand herself as a valued part of a community. Democratic discourse itself can be among

---

[16] I see no particular inherent value in participation in ownership. Broad access to ownership may be desirable for efficiency or distributional reasons, if running a commons is particularly profitable, but there is nothing distinctive to a commons there; such access is desirable to the same extent that broad access to any economic opportunity is. There are also instrumental reasons for and against broad access to ownership, which I will discuss below.

the information goods produced by an online semicommons. And sometimes justice requires that all members participate in choosing the rules by which they will be governed, particularly where the community is one that an individual has no choice but to join.

These virtues are incomparable. By talking of virtues, plural, one can discuss a community without needing to decide, say, whether more democratic participation for members justifies greater restrictions on who can be a member. Interpersonal incomparability is particularly important; one community might place more costs on authors, another might place those costs on moderators. I hope to provide a useful vocabulary for others to debate whether some such choices are better than others, but such questions are beyond this paper's scope. My goal here is to show how different forms of moderation make tradeoffs among these virtues in different ways.

C. Strategic Behavior

So far, things matters seem sunny. There are reasons to participate both in content-level communications and in infrastructure provision. There are also no direct costs of use that obviously and always outweigh the benefits of use. But the same, of course, might be said about grazing sheep on a pasture The problem is not use in itself, but overuse, or even abuse. It is not enough that some uses of the semicommons are profitable and nondestructive. It must also be the case that the semicommons is resistant to strategic behavior by participants. We are therefore led to analyze the ways in which users can impose costs on each other.

The first pair of problems involve overuse. There is *congestion*: each author's contribution requires some support from the infrastructure; overuse causes congestion. As the infrastructure is rival, this problem is straightforwardly Tragic. Readers and moderators also contribute to this potential tragedy, but to a less significant extent.[17] There is also *cacophony*: each author's contribution also raises the search costs for readers and moderators by increasing the amount of material that they must sort through.[18] This problem is also potentially Tragic; a

---

[17] There are two reasons for this difference. On the one hand, reader and moderating use fewer resources when there are few contributions from authors for them to consider. It takes less bandwidth to download the latest 5000 messages than the latest 50. On the other hand, authors consume storage space in a way that readers and moderators do not. A server can mostly forget about a reader once it has delivered to her the content she requested. The server cannot so easily discard an author's contribution.

[18] For a given reader, or for readers in general, this effect may be offset by the benefits to readers of being able to read an author's contribution, or it may not. Much depends on where the contribution falls along readers' curve of diminishing marginal returns from further options.

reader's supply of attention is limited. Spam is an example of extreme overuse causing both congestion and cacophony.[19]

Second, there are problems of *manipulation*, in which ideological moderators try to shape the information seen by readers. Where their efforts make the commons less searchable, promotes unwanted content, or hides wanted content, the moderation imposes costs on readers and authors The heterogeneity of moderators and readers means that it is hard to say anything general about when moderation is problematic, but a few common patterns are worth noting. In edit wars, moderators with conflicting ideologies wastefully competing to shape a particular output. In capture, a private owner uses her control of the infrastructure to impose her ideology as a moderator. In deception, moderators lie to the community to promote their point of view. An interesting special case of deception is sock puppetry, in which moderators create alternate personae to give the false impression of a chorus of approval for their point of view.

Third, there are problems of *weaponization*, in which authors distribute content with negative value. In manipulation, the use might reasonably be seen as legitimate by one side or another, and the problem is that subjective disagreements become destructive to the community and the platform. By contrast, in weaponization, everything is functioning properly from the Comedic point of view, save that the information outputs are bads, rather than goods. Weaponization can take the form of harassment: attacks directed at particular other users, such as vicious flaming on an email discussion list or griefing in a virtual world. It can also take the form of malice directed at other users in general; "trolls" on Slashdot post deliberately insulting and outrageous statements because they enjoy watching the frustrated reactions of others drawn into off-topic arguments. It can involve the misuse of the platform for purposes that its users find offensive, such as using an unsecured wiki to distribute pornography. And it can involve the use of the platform for purposes valued by the members but with serious negative externalities for society at large, such as a web discussion board for encouraging the assassination of doctors performing abortions.

All of these forms of strategic behavior can have an important spillover consequence: *demoralization*. Those who have negative experiences with the online semicommons may stop

---

[19] This general definition of overuse emphasizes that spam is a problem hardly confined to email. Any sufficiently advanced social technology is indistinguishable from a spam vector. In Clay Shirky's words, "Social software is stuff that gets spammed." *See* Clay Shirky, *Tags Run Amok!*, MANY-2-MANY (Feb. 1, 2005), http://many.corante.com/archives/2005/02/01/tags_run_amok.php.

participating. Demoralization is in a sense just the Tragic story from an *ex ante* point of view; those who do not expect the experiment to succeed will not take part in the first place. Readers annoyed with spam will stop reading; authors who do not expect to be read will stop writing; moderators who expect to be outvoted will stop moderating; platform owners disgusted with what they see will close off access. The demoralization problem of underprovision is the common failure condition to which all of these forms of strategic behavior tend if unchecked. To prevent such collapse, community members will try to engage in various forms of moderation and self-control.

IV. Properties of Online Semicommons

[To come.]

V. Case Studies

A. *Metafilter and Slashdot: Multiple Paths to Success*

Our tour begins with two web sites dedicated to sharing interesting links. Metafilter (a "community weblog") and Slashdot ("News for nerds. Stuff that matters.") face roughly analogous problems and have roughly analogous structure, but solve their moderation problems in remarkably different ways. Those differences richly illustrate the diversity of moderation patterns. Both sites are structured around a set of user-submitted "front page posts" (for Metafilter) or "stories" (for Slashdot): links to something interesting on the web, along with a few sentences of description. Each story has an accompanying discussion page, on which users post their own commentary and conversation sparked by the initial link. Slashdot has about 25 stories a day, typically with a few hundred comments each, and about 150,000 registered users. Metafilter also has about 25 stories daily, typically with 20-100 comments each, and about 50,000 registered users. More people read both sites than post to them; Metafilter sees about 3 million unique IP addresses s a month; Slashdot, about 7 million. Both communities are mid-sized by Internet standards, and orders of magnitude larger than the size at which one normally expects cooperative norms to start breaking down.

Slashdot filters all story submissions through a small (half a dozen) set of moderators, who choose which stories will appear. The number of submitted stories is surprisingly small—a couple of hundred a day—which makes this human centralization tenable. Comments, on the other hand, are wide open. Both registered users and "Anonymous Cowards" can post comments immediately. The need to associate a comment with a story provides a first level of structure,

and comments can be threaded to create conversations. Registered users can "moderate" other users' comments, giving them a +1 or -1 ranking for various reasons, such as "interesting," "insightful," "offtopic," or "troll." These scores are summed on a range from -1 to 5; readers can then choose to see only those comments above a certain threshold. Slashdot quickly discovered that moderation could also be abused, and has instituted increasingly baroque systems to watch the watchmen. In "meta-moderation," registered users judge whether randomly-selected moderation decisions were justified or not. These meta-moderation decisions then become feedback into a "karma" system under which only those moderators who use their powers for good rather than evil (as judged by a complex but open-source computer algorithm) are allowed to moderate in the future.

The moderation/metamoderation/karma system is the core of Slashdot's self-regulation, but there are other systems at play. IP addresses that try to engage in denial-of-service attacks or post malicious comments will be banned. There is an experimental user-supplied tagging system. Comments with malformed HTML (which could create security holes or rendering problems) are blocked by software and will be removed on sight by the site admns if discovered. The site shows advertising, but readers can purchase subscription that give them a thousand ad-free pageviews for $5. Slashdot's norms are particularly interesting. The site has an explicit technolibertarian philosophy; its administrators do not delete posts for reasons of content and they have fostered a climate of free-wheeling high-intensity exchange, in which tempers frequently flare and insults fly. The site has had an extensive history of deliberate misbehavior; users would attempt to score a "first post" by jumping on a newly-posted story, to horrify readers by displaying disgusting images such as the infamous goatse, or to troll other users. In fact, flagrant misbehavior of this sort is still regularly attempted but is now routinely moderated down and out of sight (most users set their comment filters to 0 or above, so that the -1 posts simply disappear from view). Slashdot discussions still range from the sophisticated to the sophomoric, but even a casual browse reveals that the moderation system is effective in screening out most of the true dross. Indeed, as a long-time Slashdot reader, I can report that first-post and goatse attacks are far less common now than at the turn of the millennium; users seem to have learned that these techniques are generally futile.

I would not say that Metafilter could not be more different from Slashdot—it could—but the differences are nonetheless striking. Where Slashdot membership is open to anyone

immediately; Metafilter requires a $5 fee to sign up. (During times of crisis, such as when a group of outsiders is flooding in, new signups may be disabled.) Where Slashdot filters posts through moderators, Metafilter lets anyone with an account create an FPP. Comments are organized by FPP, but otherwise are unthreaded. Users can mark their favorite FPPs and comments, and also flag FPPs and comments for removal. Favorites are displayed publicly, while flagged posts are shown only to the site administrators. The three administrators delete posts and comments for various reasons, usually with a short standard explanation, and paying special attention to repeatedly-flagged ones. There is a norm-setting asymmetry here; positive contributions are praised and made more visible, while negative ones are hidden. Matt Haughey, the site's founder and lead administrator, also maintains a sideblog on the front page of "new and noteworthy posts" that also highlights particularly excellent contributions. Users who behave disruptively will draw disapproving comments from others users, then gentle emailed reminders from Haughey, then more serious warnings, and finally an outright ban. The site is advertising-supported; Haughey has cycled through a number of different advertising systems,[20] all of them unobtrusive.

Haughey and his right-hand co-admin, Jessamyn West, are absolutely certain that the secret to Metafilter's success is generating positive, cooperative community spirit—in my terminology, norm-setting. They think of all of their moderation actions in terms of how they will affect group norms. Thus, they are heavy and visible site users themselves (Haughey credits the site's initial takeoff to the weeks during which he was the primary author of FPPs and active in all discussion threads, setting an example for others) and heavy email correspondents, offering pats on the back and gentle reproaches. Offshoot sites provide additional opportunities for users to blow off steam, provide feedback on the site, and learn about each other: particularly MetaTalk, where users discuss Metafilter itself. They are willing to explain any action to death, try very hard to keep their own emotions out of their actions, and move quickly to push back against actions that threaten shared norms of cooperation. Partly as a result, Metafilter has a solid (if constantly changing) core of about 300 heavy users who reinforce norms of basic respectfulness and self-restraint, and a longer tail of thousands more who participate on these

[20] He first allowed MetaFilter members to purchase small textual ads, then switched to Google's textual ad system, and then joined the boutique Federated Media banner advertising network, accepting only small banners in the right-hand column. He also arranges one-off deals; he ran a one-day banner ad for the Serious Eats blog in exchange for a gourmet dinner.

largely positive terms. There have been a string of crises over the years—real disputes among camps of members, invasions by crowds unfamiliar with Metafilter but angered by something posted there, near misses with upset lawyers, and occasional harassment—but the site has successfully weathered multiple scale transitions with a basic sense of intellectual ferment roughly intact.

Metafilter and Slashdot's experiences have some important similarities. Both are run by administrators who are deeply involved in day-to-day conversations on the site, who rely both on their own participation and extensive statistical instrumentation to keep in view what is happening. Neither is at all a democracy in a formal sense, in that the administrators on both set policy unilaterally. In West's words, "At the end of the day, someone always has root; ignore this at your peril." Both now rely on core groups of participants; Slashdot has a smallish core of regular story contributors, while Metafilter has its version of the 300. Both produce a diet of links and a much larger flow of conversations, both of which are thrown open to the wider world. But the richness of moderation is such that they achieve their success using remarkably different patterns. Slashdot's moderation/metamoderation system is a complex system of aggregating user feedback on other users in an algorithmic way; Metafilter's "system" is really a rich set of norms, in-jokes, customs, habits, and styles. Slashdot is stable because most users want it to work and because its algorithms have been well-tailored to provide a strong lock-in that prevents a few users from disrupting the organizational machinery. Metafilter is stable because it has a critical mass of dedicated users who apply lesser social sanctions, who are backed up by conscientious and active administrators who can apply stronger sanctions as needed.

B. USENET and Email: Not All Moderation Succeeds

USENET newsgroups (started in 1979) and email (1982, based on protocols dating back to 1971) are both old technologies, by Internet standards. They have roughly similar overall structure; both are systems that allow users on computers across the Internet to communicate through a peer-to-peer process of message exchange with local storage. Email prospered on the massive scale-up of the 1990s Internet boom; USENET was overwhelmed by that same boom. Their divergent paths illustrate how the choice among moderation patterns can doom or save an online community.

USENET is a distributed set of message boards, with a peer-to-peer protocol by which different sites exchange new messages. Each server talks only to a few others, but most

USENET servers are linked together so that any given message will eventually be propagated to all servers in the network. In 1987, participating sites agreed to a "Great Renaming," establishing a canonical hierarchy of topical newsgroups (such as rec.pets.cats, sci.math, and alt.fan.karl-malden.nose); a given message is posted to one or more newsgroups. Some groups require the approval of a moderator to post a message; others do not. Posts can be deleted by the author; these "cancel" messages are easily forged, so site administrators often set their own policies on whether to honor cancel requests. Many users use "killfiles": lists of other users whose posts will be hidden from the killfiler. Creation and deletion of newsgroups normally goes through a deliberative voting process coordinated by a centralized board that publishes a canonical list of approved newsgroups. In the alt.- portion of the hierarchy, however, site administrators simply decide whether or not to add a new group, and other administrators decide whether to follow their lead.

This abbreviated description of the system reveals many familiar organizational techniques, including annotation (choice of newsgroup), filtration (killfiles), and deletion (cancels of others' posts). The hierarchy system has centralizing democratic procedures almost hardwired into it, but also respects decentralized values with the less-constrained alt.- hierarchy. The peer-to-peer setup gives it good scaling properties, since sites bear much of the costs of hosting and communicating created by their usage. The division into different newsgroups also generated strong group community norms within different newsgroups; people drawn together by shared interests have often developed strong community feeling.

For all of these virtues, USENET's moderation patterns did not deal well with the flood of new users who cam online in the 1990s. In 1993, AOL began offering its subscribers access to USENET newsgroups, an event known as the "Eternal September"—a neverending stream of new users as unfamiliar with USENET's norms as the annual crop of college first-years had been. In 1994, the commercial spammers arrived, starting fittingly enough with a pair of lawyers who cross-posted an advertisement for their immigration services to thousands of newsgroups. Massive cross-posting, combined with a lack of respect for norms of self-restraint, became endemic. Griefers made both targeted attacks on particular newsgroups and general attacks on hundreds at once; when an unsophisticated user would try to respond to insist that they leave, she would often simply end up cross-posting it herself, adding to the chaos. Vigilante USENET users began automating cancel requests to counter automated cross-posts, but the damage to

many previously healthy newsgroups was fatal. As web-based discussion boards and third-party hosted mailing lists become increasingly reasonable alternatives, many of the conversations that had taken place on USENET shifted elsewhere or simply died. USENET itself is not dead—one can still go to Google Groups or Giganews and participate in many ongoing conversations in groups with strong norms that allowed them to weather the storm—but it has nowhere near the relative importance that it once did to the life of the Internet. USENET fell victim to the moderation equivalent of the Peter Principle; it grew until its moderation mechanisms could no longer cope with the new scale of disruptive behavior, and there it stopped.

The email system did not stop growing in the mid-1990s, even though it, too, was deluged with spam and (to a lesser extent) with denial-of-service flood attacks. USENET's Achilles heel was that it had centralized organizational patterns but no effective system by which users could pool their efforts to defend that organization from large-scale attack. Each newsgroup spanned all servers; each server hosted all newsgroups. Unlike USENET's NNTP, SMTP is not a replication protocol, but a transfer protocol. Emails go to their recipients, full stop. There are no email equivalents to newsgroups—coordinated entities that all users see in a substantially identical form. Each email server is a dedicated piece of infrastructure designed to enable incoming and outgoing email for its own users. I can install a spam filter without disrupting any email except that to and from users on my piece of the network. This difference has provided for greater flexibility in experimentation with local anti-spam policies. Spam remains a serious, expensive problem, but email remains usable and valuable. The network of regular email users continues to grow.

The tradeoffs between freedom and control in the email system are fascinating. Email scores highly on many virtues of openness. Access to email is exceedingly broad, and that access is close to free for users. Once one has an Internet connection, one can usually get free email from one's ISP or ad-supported email from a webmail provider. Individual users have substantial control over their own spam filters, a little say in their ISP's spam filters, and almost no say in any global properties of the email system (which has little in the way of formal governance). Individual-to-individual email is mostly uncensored and uncensorable by other users. But all of these virtues come at the price of some specific design tradeoffs. Email is deeply non-public; a great email message can only be shared with many people through successive forwarding, thereby foregoing some productivity. The decentralized system that

makes spam locally filterable also makes it hard to stamp out spam on the sending side, creating large back-end costs and moderate costs for users who must tend to their spam filters. The inconsistent and decentralized system of anti-spam enforcement also creates unaccountable, sometimes undetectable drop-outs in email connection—messages sometimes are thrown away without warning and for reasons that may be hard to fathom. Every large mailing list moderator has horror stories. There are regular calls to make email more accountable—making senders pay to send messages, creating stronger authentication, or establishing central anti-spam systems. Centralization of these functions would raise difficult democratic issues of participation and accountability, would threaten users' abilities to experiment different local spam policies, and would threaten the live-and-let-live approach that allows users with violently different beliefs all to use email.

C. Wikipedia: Costs and Benefits

Wikipedia is a favorite example of Comedic theorists. Its rapid growth and surprising reliability make it a powerful proof-by-example of the power of large-scale Comedic sharing. I do not plan to rehash what has already been written about Wikipedia's sudden success. Instead, I would like to focus on an underappreciated aspect of Wikipedia's moderation policies: their dynamic evolution. Wikipedia has thrived in part because its moderation policies have adapted to changing challenges.

The founding moment of Wikipedia is usually considered to be Jimmy Wales and Larry Sanger's January 2001 decision to augment the peer-reviewed Nupedia with a wiki open to anyone to edit. This move shifted from a system with substantial exclusion barriers for author to one without, and also to one with a much less (implicitly) expensive contribution process. This new process was much more friendly to rapid large-scale editing, and catalyzed rapid growth both in the number of articles and in the number of contributors. (Wikipedia relies on synthesis in a deep way, allowing editors to make incremental improvements on previous contributions.) The success of these early efforts, in turn, had a useful norm-setting effect in drawing more contributors in. It was, in short, a virtuous cycle. And at first, page vandals could be kept in check simply by the regular efforts of conscientious editors. The subsequent history of Wikipedia is also a history of innovative responses to threats. Here are three interesting examples:

Protection and Blocking: Wikipedia's growth has made it attractive both to vandals and to ideologues. Users who engaged in repeated or large-scale vandalism, who take part in vicious edit wars and will not back off to discussion when asked to, who target other users, or who flout other important community polices are now subject to IP-addressed based blocking, a form of ex post exclusion. This policy is exercised in tandem with protection, under which controversial or important pages at a high risk of vandalism or ideological edit wars are "protected" with varying degrees of intensity from edits. Full protection prevents everyone other than administrators from editing a page; semi-protection prevents new users from editing pages. These are forms of ex ante deletion. Protection is regularly used not merely to prevent norm-defying users from making changes, but also to reassert norms without aliening users by establishing "cooling off" periods.

Dispute Resolution: Wikipedia's reliance on centralized synthesis gives it a classic unavoidable problem common to such schemes: resolving disputes over which view will prevail. To an extent that should not be underestimated, Wikipedia often simply ducks these issues. Many Wikipedia entries adopt a measured "on the one hand . . . on the other hand . . ." tone towards many questions, legitimating some quite questionable views in an excess of caution. This response, as frustrating as it can sometimes be in a would-be authoritative encyclopedia, is an understandable technique of community maintenance that avoids alienating participants by rejecting their views. Wikipedia also uses a confusingly wide array of dispute-resolution techniques, including taking votes on article-deletion requests and a system of tiers of review for policy decisions. These mechanisms are largely self-generated, an outpouring of cooperative creativity that would make Ostrom proud.

Participation-Based Identity: Having abandoned Nupedia's system pre-credentialling, Wikipedia is faced with difficult identity-related challenges. Matters such as page deletion are conducted by vote among those interested, making sock puppetry attractive to those who would win a dispute. More generally, in a large community where any user has the technical power to alter the outputs, some system by which users can sort out who else to trust in case of dispute is essential.[21] This system of trust is social; long-time Wikipedians who have made many edits,

[21] Another important prerequisite for stability is part of any wiki software: keeping page histories so that any mistaken edit can be undone, or "reverted." No user actually has the power to delete previous

participated in many discussions, and been praised by many other Wikipedians have greater socially-granted authority. This measure of community respect is used in determining which Wikipedians will become administrators with greater code-based powers. It is also used as a sign of authority in disputes all across the site. Long-time Wikipedians know about each other and will often stick together to reassert Wikipedia norms and internal operating rules. This social coherence has provided legitimacy for many organizational projects and procedural standards that give Wikipedia and Wikipedia editing a somewhat predictable structure.

These mechanisms all have costs. Protection inhibits editing of many pages of great importance; disputed topics are sometimes simply frozen as is for the sake of stability. The dispute resolution systems are often ad hoc and confusing. And the informal credentialing system favors long-time contributors substantially invested in Wikipedia. All three mechanisms have played roles in helping Wikipedia grow and maintain coherence against both directed attacks and emergent chaos. Especially taken together, however, they create an increasing risk of capture—newcomers are presented with a set of complicated and inconsistent community rules, which they are then penalized for not knowing. Prizing community contribution over intellectual contribution is a natural pattern to adopt. But for an encyclopedia, it has substantial risks—including an open disdain for expertise, a sometimes comical lack of interest in contributor honesty (especially when dealing with non-Wikipedians), and hostility to new users. The communal social experience of Wikipedians both sustains and threatens Wikipedia as a knowledge-generating institution; it is both an instrument of freedom and of control; it both deters and constitutes strategic behavior. Wikipedia's ability to develop these institutions is a testament to the power of its deliberatively democratic processes and the responsiveness of those with control over the code to the needs of the community. With them, it has managed multiple scale transitions and several crises of credibility. But these institutions themselves are responsive to current problems, and Wikipedia's continued growth will only introduce new ones, even leaving aside the danger of hostility to outsiders. The most positive indicator for Wikipedia's ability to continue its success is not just that its moderation patterns currently work, but that it has developed moderation patterns dynamically in response to past challenges.

VI. Conclusion

---

contributions. (Deleting entire pages is a difficult border case because the deletion removes the page history. Unsurprisingly, Wikipedia has special policies on point.)

[To come.]