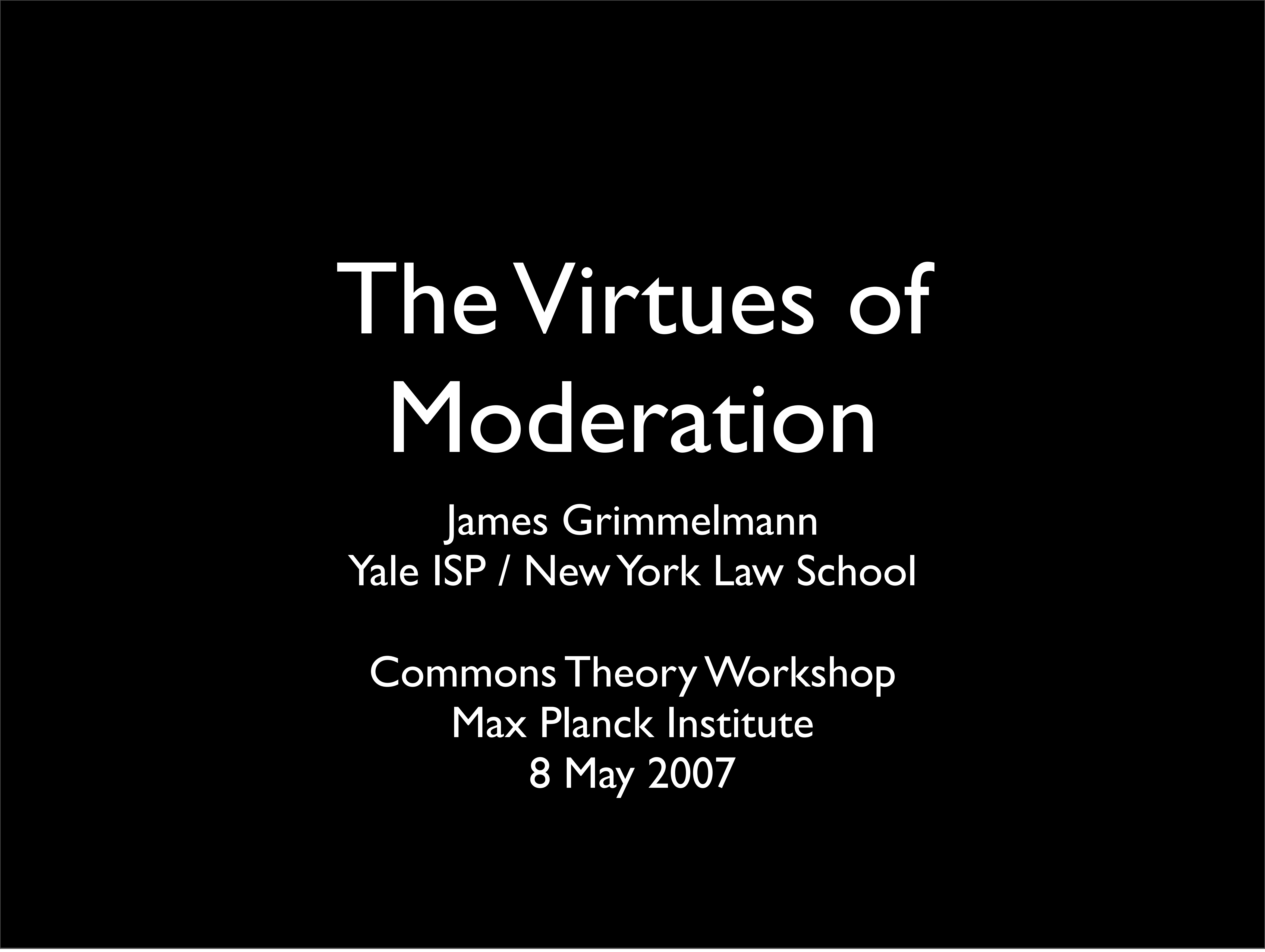


The Virtues of Moderation

James Grimmelmann
Yale ISP / New York Law School

Commons Theory Workshop
Max Planck Institute
8 May 2007

Recent scholarship, especially as synthesized in Yochai Benkler' work, has made strong claims that a commons in information goods can promote massive collaboration. This collaboration is economically attractive, because it can involve efficiencies of distributed production without incurring the costs associated with exclusive rights. It's also attractive from an access to knowledge perspective, because it excludes no one (or at least, far fewer) from the information, and enables larger communities to participate in the production process.

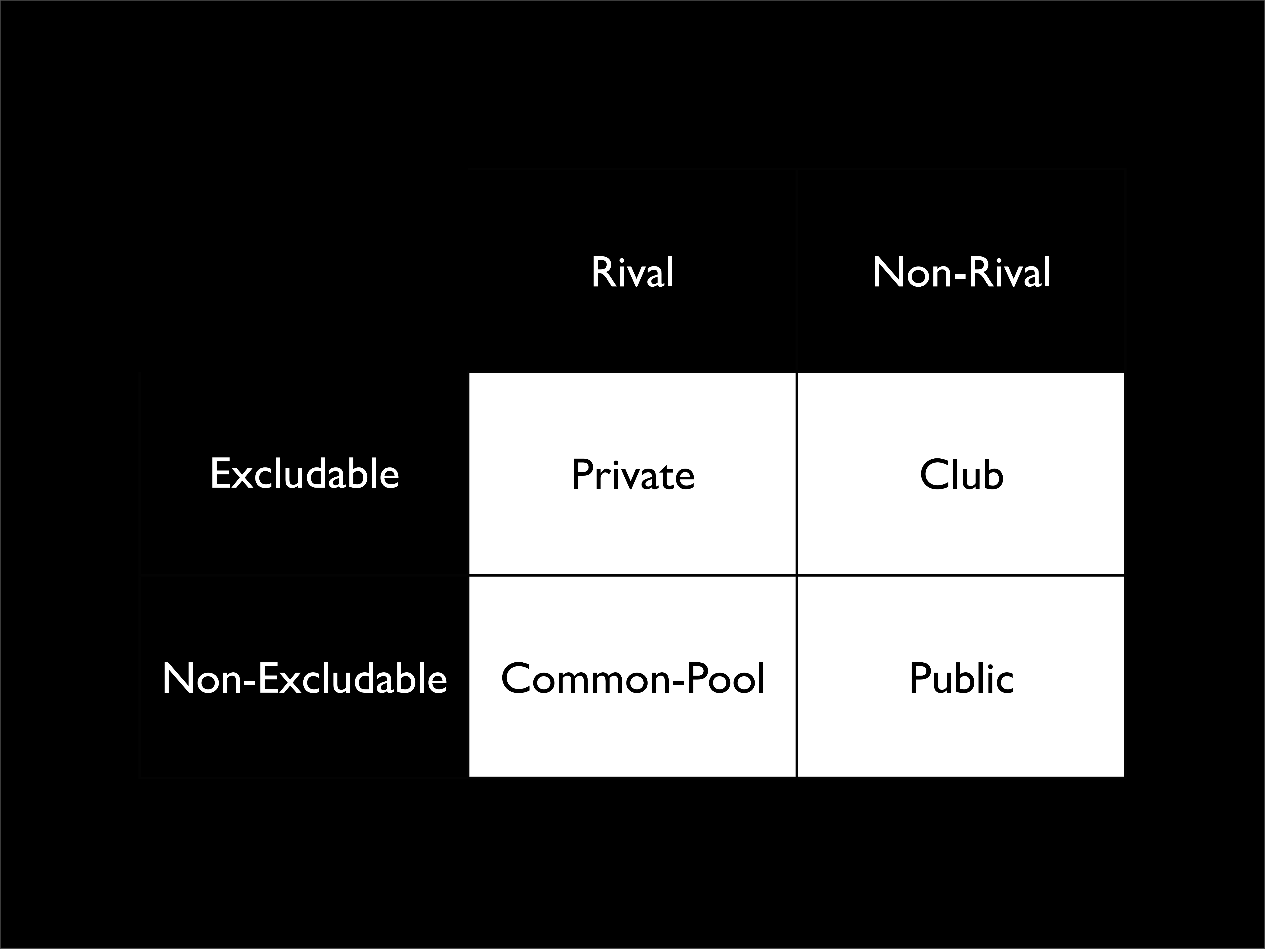
Broadly speaking, this story requires us to revise conventional wisdom in two ways. First, there is the argument that without the financial incentives available through exclusive rights, no one in their her right mind would contribute to an information commons. That argument has been fairly well-refuted, both by abundant evidence that millions of people (all apparently in their right minds) do contribute to such commons and by more nuanced explanations of why they do so. A mix of indirect economic benefits with altruism, reciprocity, and other social motivations motivate contributors to open source software and other information commons.

Communities balance freedom and control through moderation.

The second skeptical concern is that even well-intentioned and well-motivated collaborators must nonetheless face the organizational problems that bedevil firms, governments, and other institutions that must integrate the perspectives and contributions of multiple participants. Here, we have seen many useful observations and the occasional note towards a theory of commons collaboration, but nothing as general. The door remains open for skeptics to worry that unmanaged information commons will prove unsustainable as Gresham's law reasserts itself.

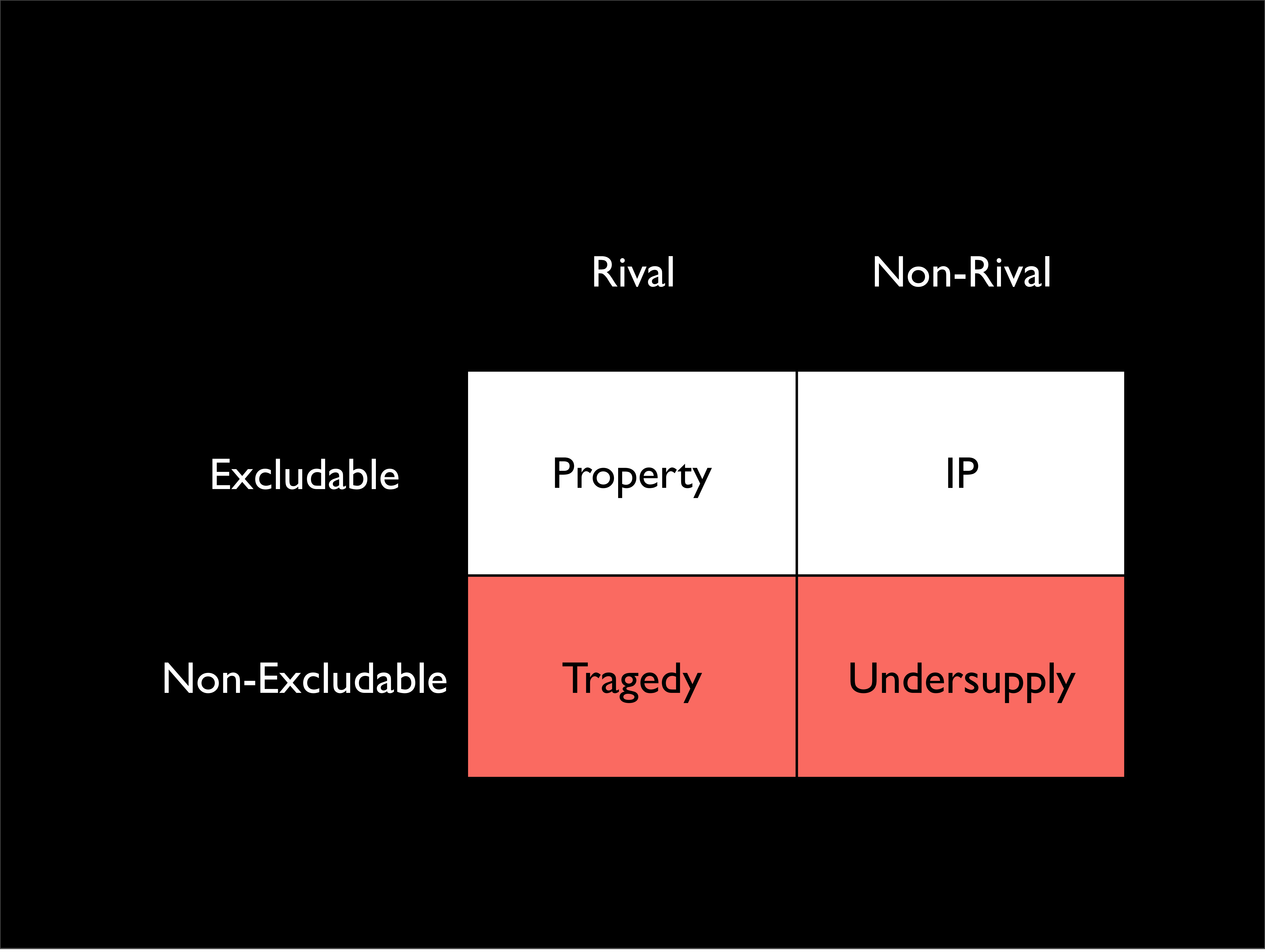
Today, I'm going to argue that the skeptics have a point, but that we have reason for optimism. That reason is moderation. I'm playing off of two meanings of "moderation" here. On the one hand, a "moderator" keeps order during a debate. A good moderator can keep hostile factions from each others' throats while helping them reach agreement on many points. Moderation in this sense is a kind of community management, both imposed and self-generated. On the other hand, "moderation" is the avoidance of extremes. A successful community is neither too free nor too controlled. Good moderation involves closing off enough features that participants don't take advantage of others' good will, while leaving open those features that catalyze collaboration.

I'm going to talk about the patterns of moderation that many online communities use. I'll explain how those patterns involve choices and compromises. And I'll argue that, for a number of reasons, we should be optimistic that many online communities will be able to find fairly stable patterns of moderation that preserve substantial productive freedoms. There are controls in such communities, but not controls that compromise the nature of the commons itself.



	Rival	Non-Rival
Excludable	Private	Club
Non-Excludable	Common-Pool	Public

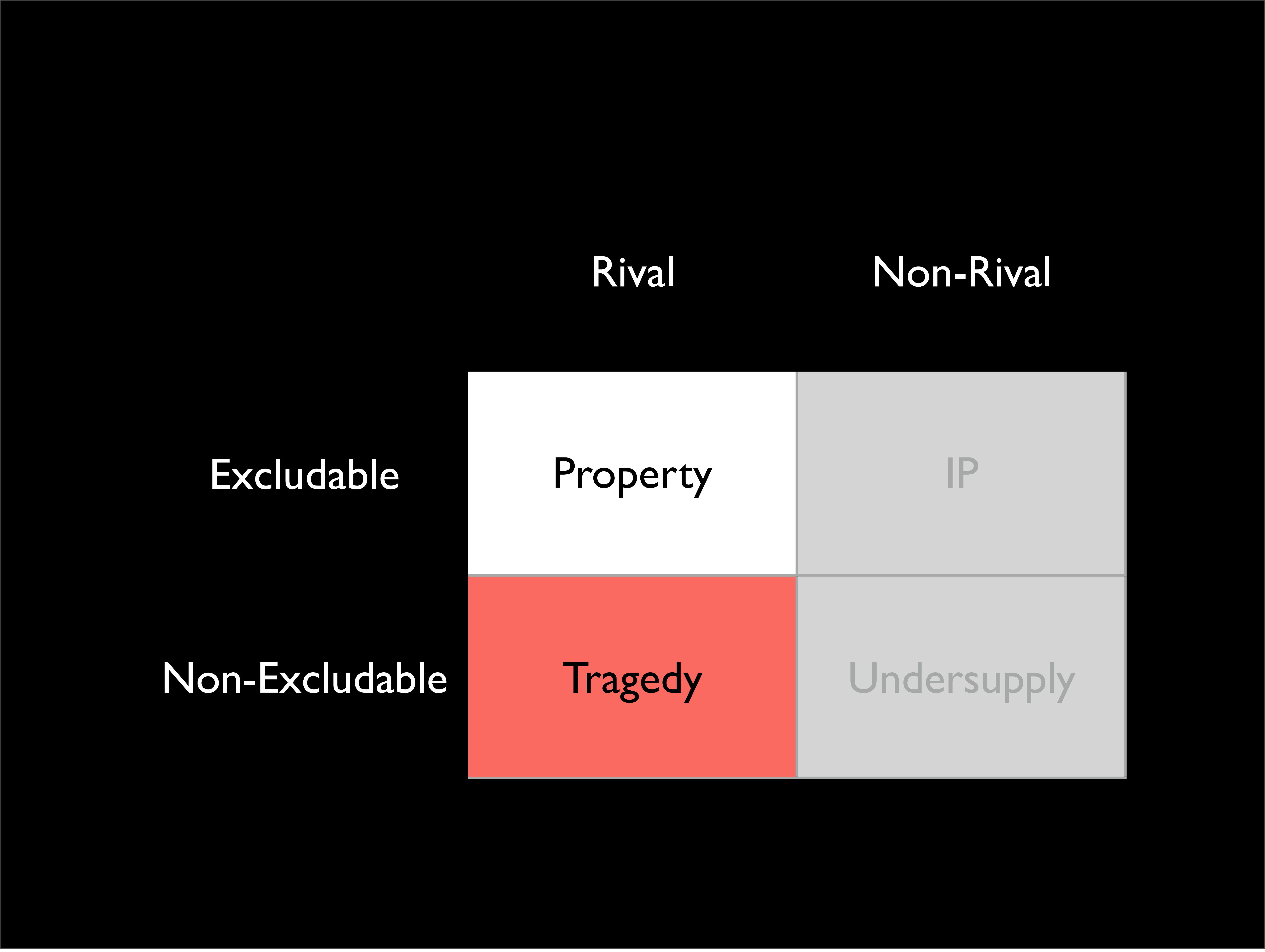
Let’s start with a little public–good economics to show the tension. Conventionally, goods are public or private. Public goods are nonrival and nonexcludable; private goods are rival and excludable.



	Rival	Non-Rival
Excludable	Property	IP
Non-Excludable	Tragedy	Undersupply

On this conventional view, it’s non-excludability that causes problems. What we can keep others from using, we can create pricing regimes in.

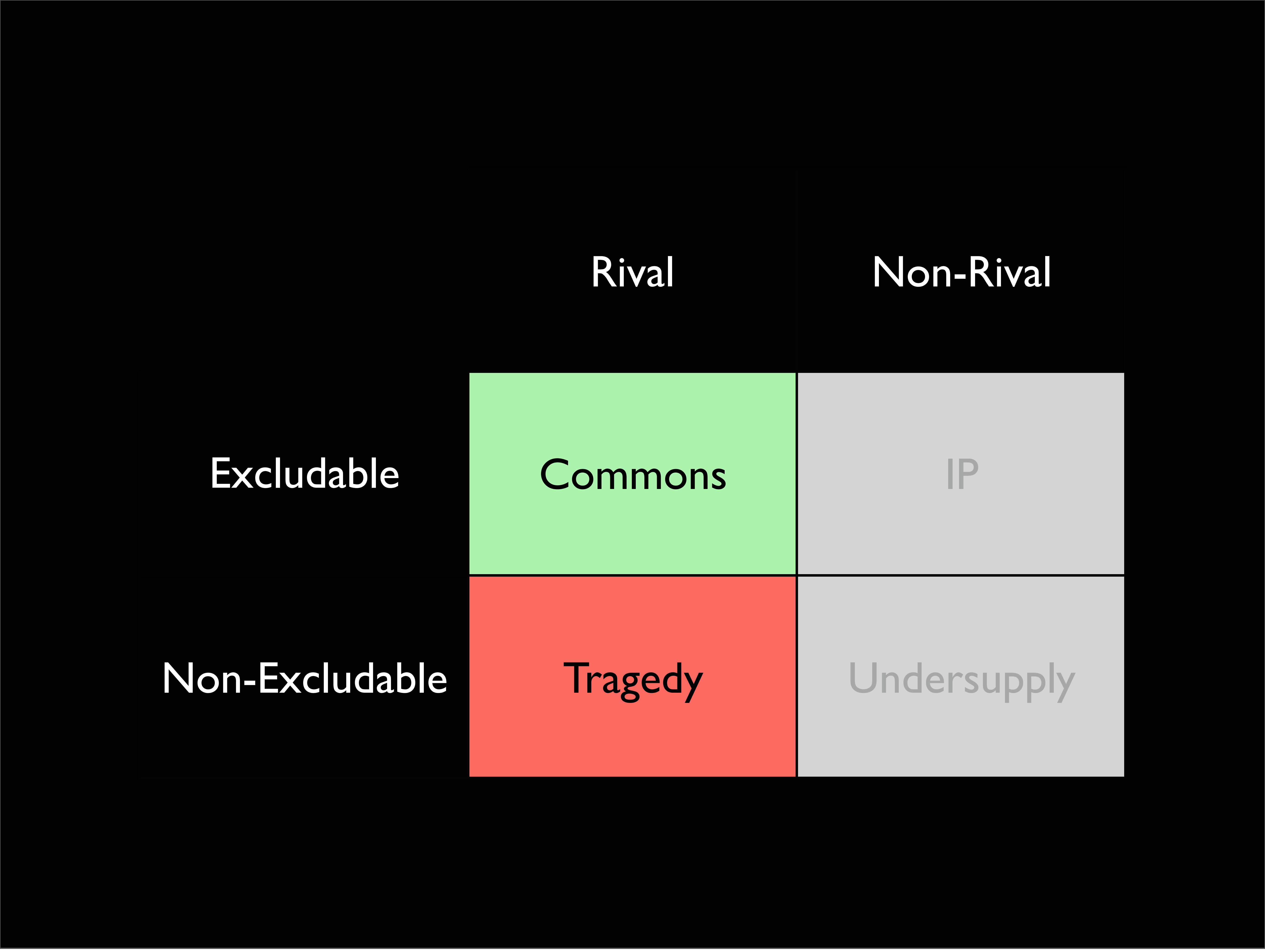
Commons theory challenges this view in two ways.



	Rival	Non-Rival
Excludable	Property	IP
Non-Excludable	Tragedy	Undersupply

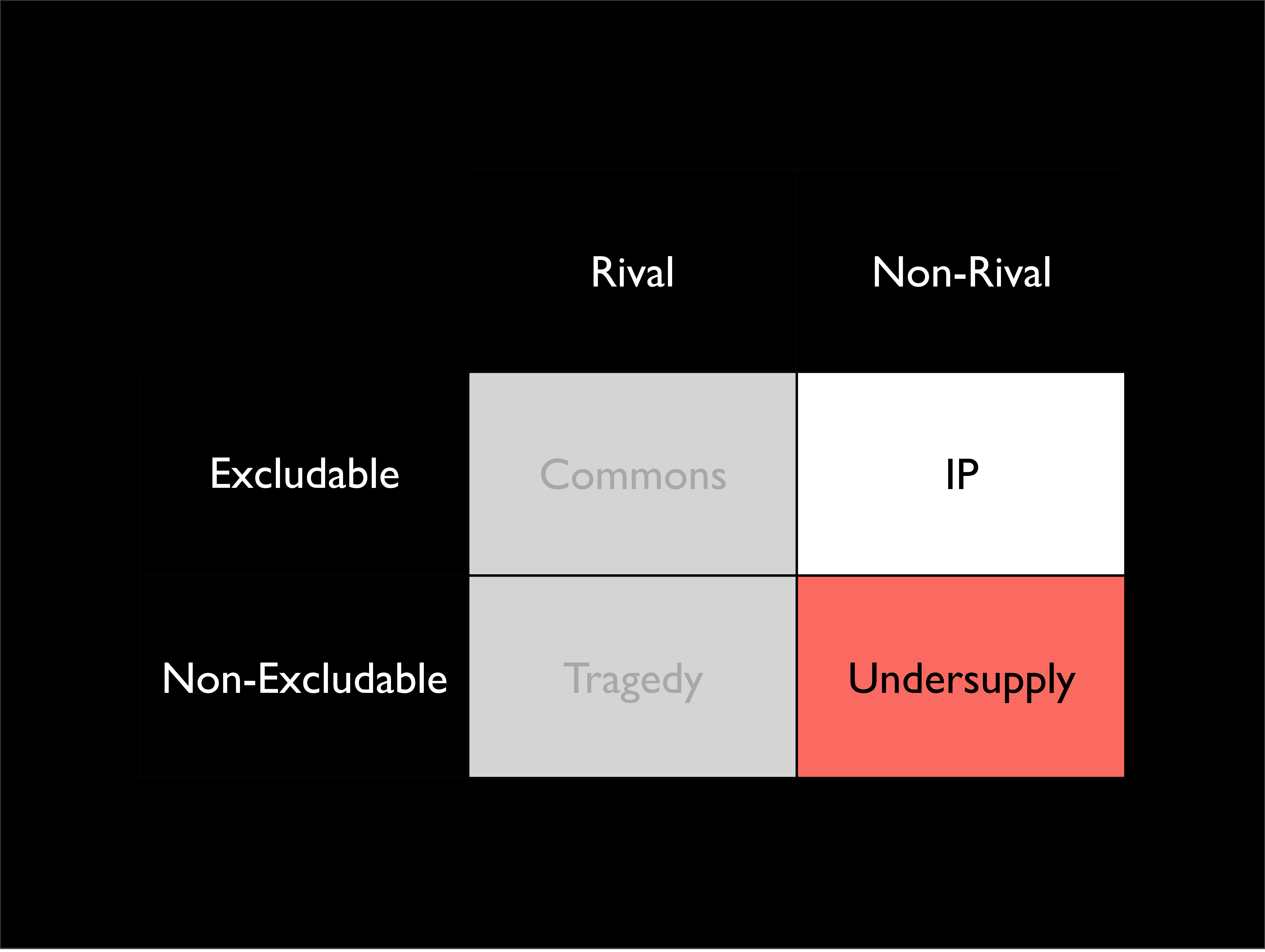
Let’s start with rival goods.

Here, the conventional theory tells us that an exclusion-based system leads to efficient allocation. Without exclusion, a tragedy of the commons results from wasteful overuse. Property rights and prices -- or perhaps, government regulation -- are needed to keep use to appropriate levels.



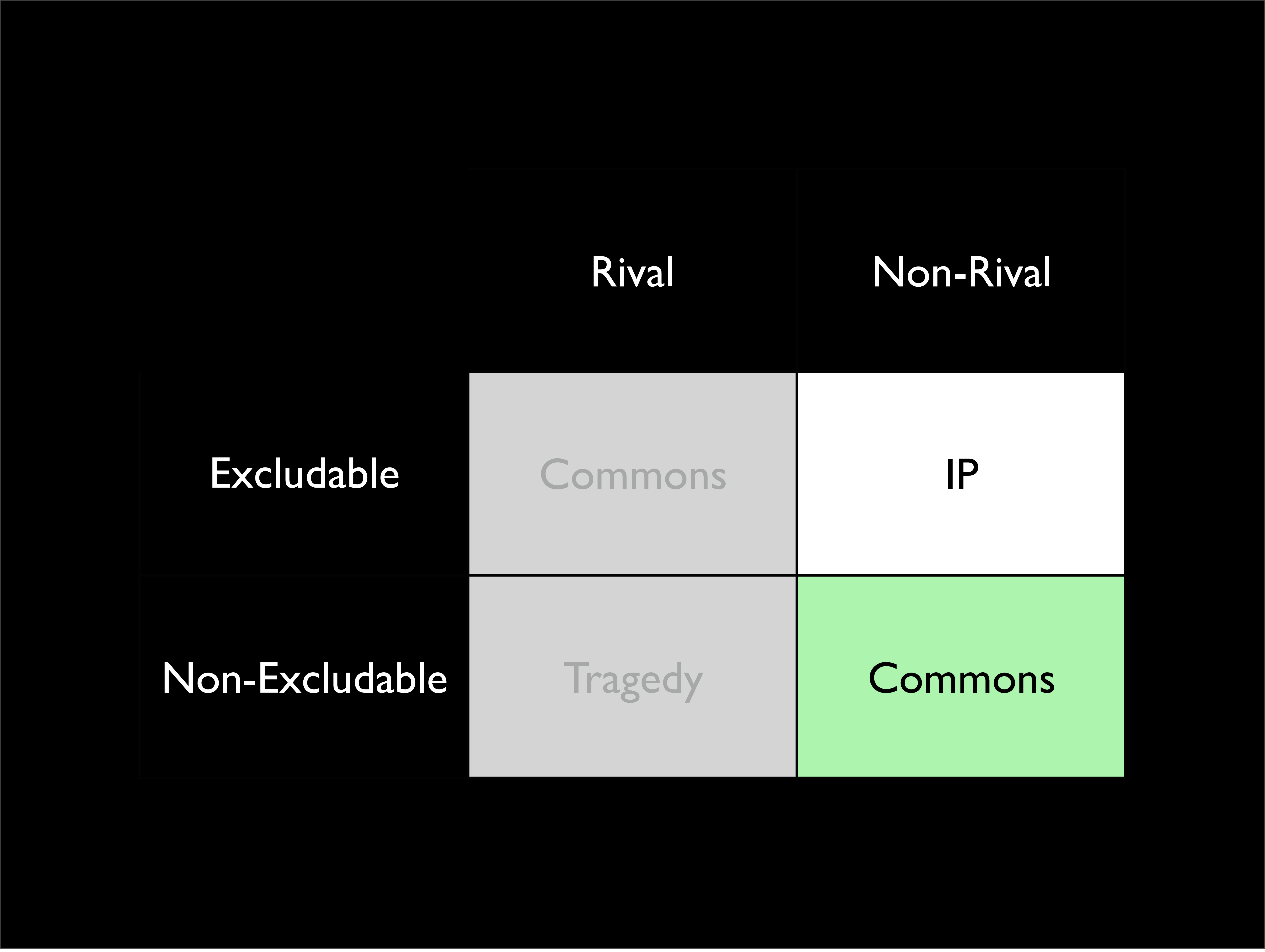
	Rival	Non-Rival
Excludable	Commons	IP
Non-Excludable	Tragedy	Undersupply

We know now that that story is incomplete. Top-down rights and regulations are not the only way to create the necessary excludability. Bottom-up self-created and self-enforced community systems of common ownership and management can also prevent wasteful overuse. The common-pool resource literature tells us that groups with well-defined boundaries, graduated sanctions, and good fora for communication can produce stable institutions that regulate use to sustainable levels. I call this the Tragic story; it explains how common ownership can avoid the tragedy of the commons.



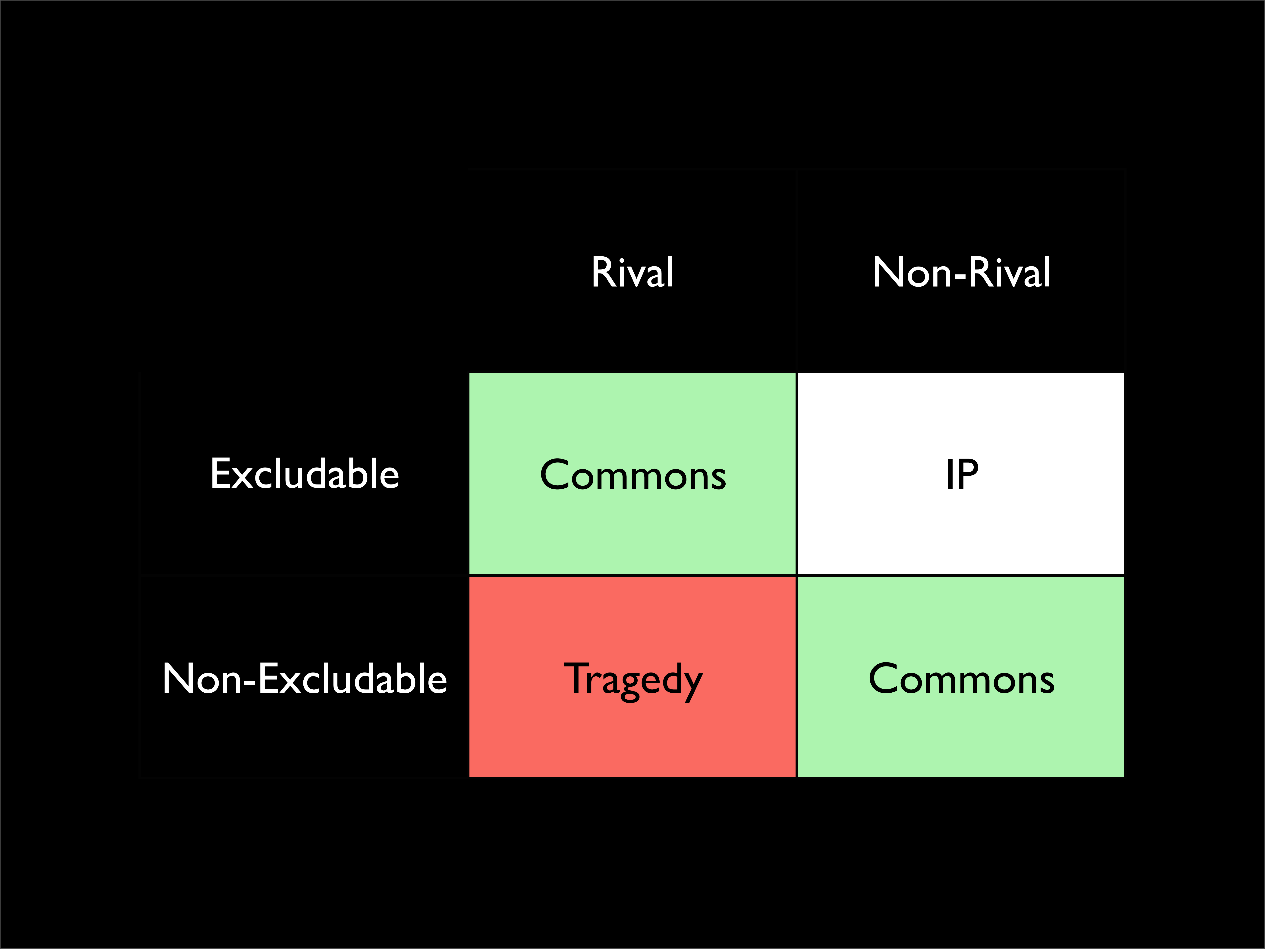
	Rival	Non-Rival
Excludable	Commons	IP
Non-Excludable	Tragedy	Undersupply

On the non-rival side, things are somewhat different. Here in the realm of ideas and intellectual property, conventional theory claims that a pricing system (or direct government provision) has its problems, but still often beats the alternative. Intellectual property can lock up information goods so that too few have access to them, but this sacrifice is a necessary one, since no one would have an incentive to create those goods in the first place.



	Rival	Non-Rival
Excludable	Commons	IP
Non-Excludable	Tragedy	Commons

Here, however, commons theory argues that there is once again a better way. The seemingly intractable production problem is in fact tractable. People are natural information producers, demand creates its own supply, and combining the creativity of huge numbers of individuals can produce all the information we ever need, and more.   Once the information exists, the best thing to do is share it as widely as possible -- we get more creativity from connecting authors to audiences and to previous authors than we do by offering them exclusive rights. I call this story the Comedic one; it explains how a commons can catalyze collaboration on a vast scale.



	Rival	Non-Rival
Excludable	Commons	IP
Non-Excludable	Tragedy	Commons

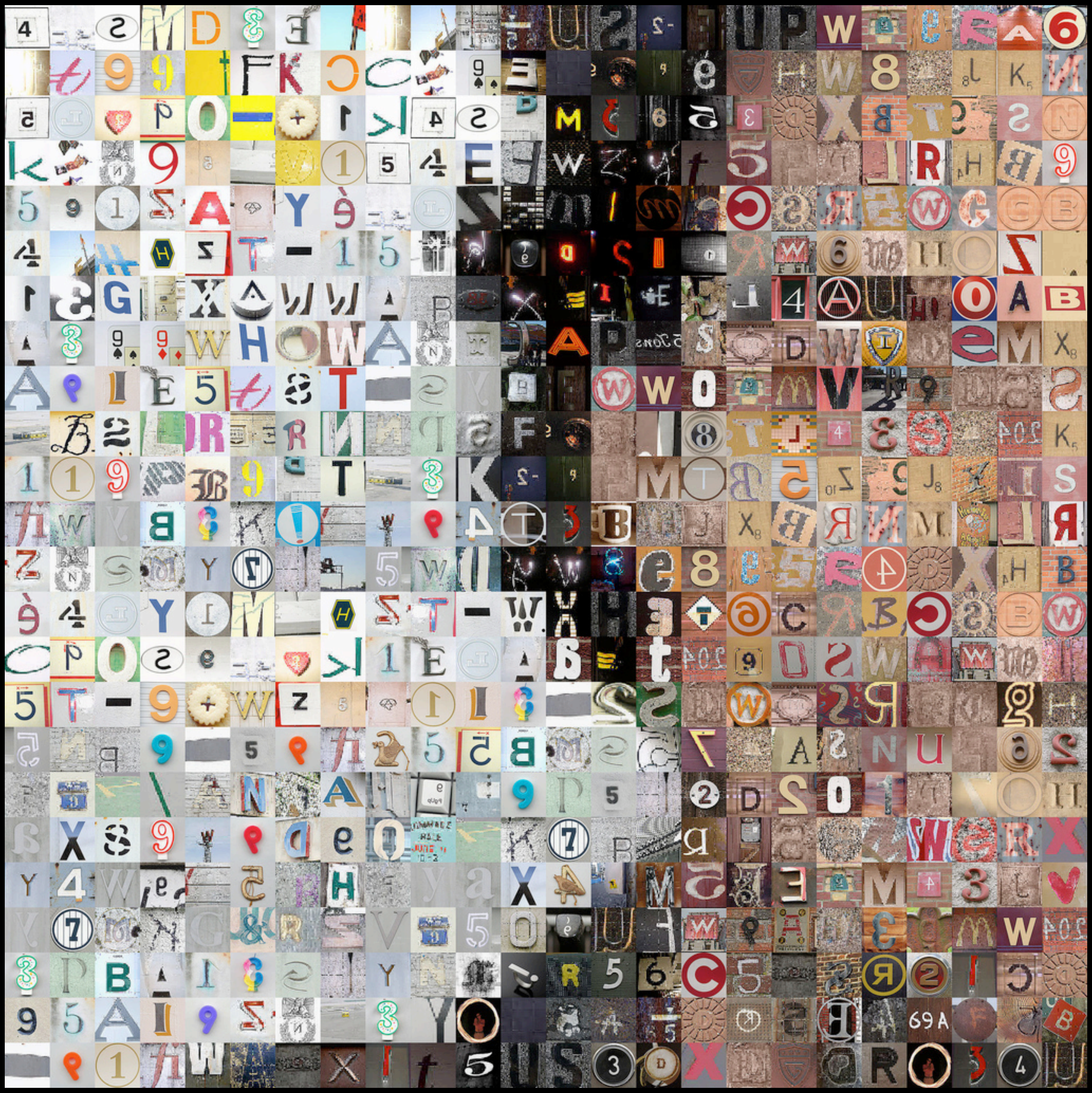
The Tragic and Comedic stories both counsel against private property rights. But they have interestingly different things to say about excludability. The Tragic story says to embrace excludability to prevent waste; the “commons” is just another institution for generating excludability. The Comedic story says to foreswear excludability and embrace the very open access that the Tragic story considers, well, tragic.

The key to absolutely everything I’m going to say here is that both stories are right, simultaneously. When we talk about particular online communities (rather than talking about a commons in information in general), these communities have some aspects that are rival and subject to Tragic effects and some aspects that are non-rival and subject to Comedic effects. To thrive, they need to provide excludability in some forms and not in others. My paper, to repeat, is about this balance between control and freedom.

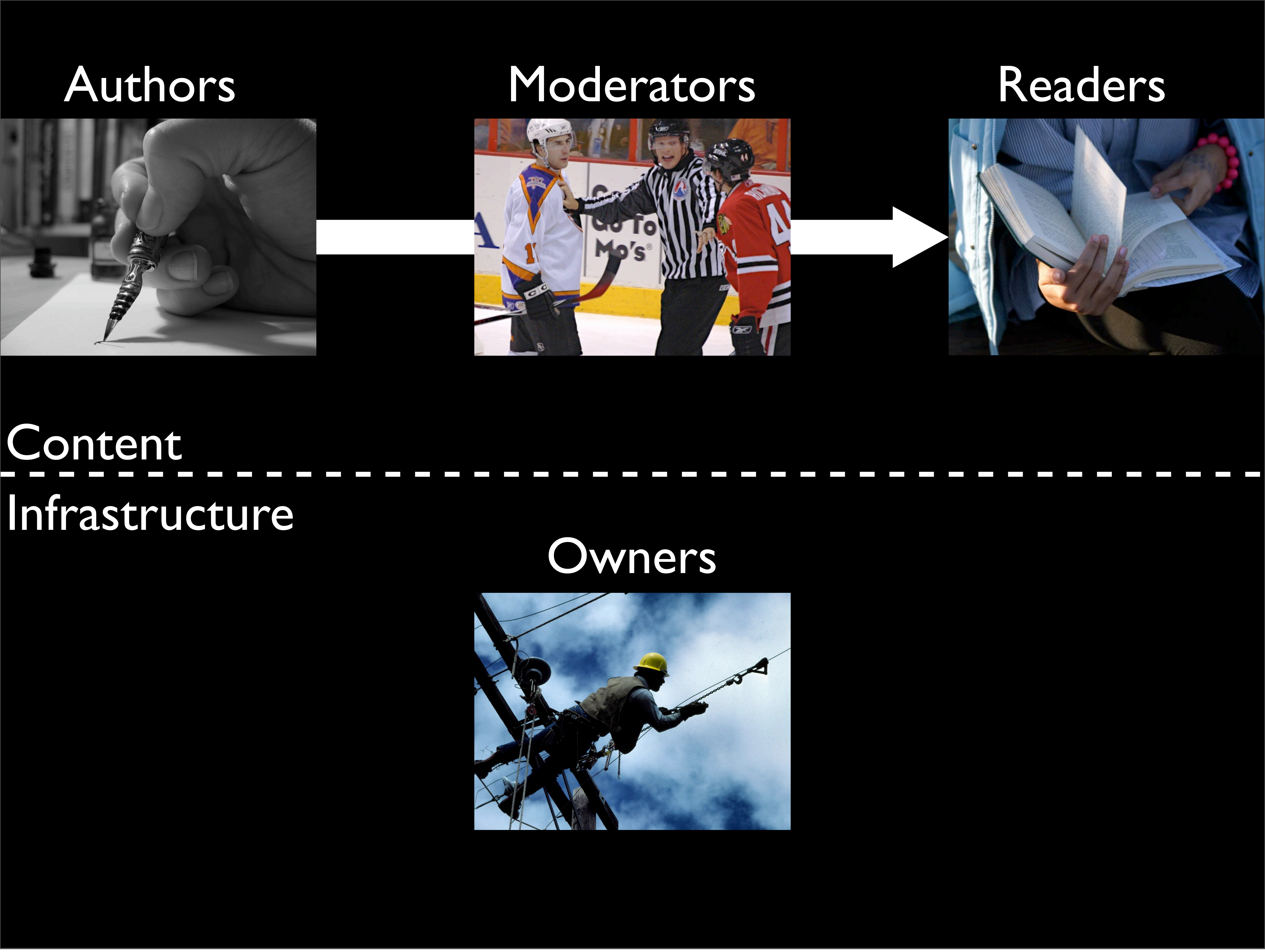
Offline semicommons

The framework I'm using comes from Henry Smith's work on semicommons resources. His theory starts from a study of the open-field system. The underlying land was held privately in strips by farmers, but during some seasons held common for grazing by sheep. This mixture of regimes had symbiotic benefits: simultaneous use, fertilization of crops, and fodder for the sheep. But it also had costs, including risks that shepherds would selectively graze on particular farmers' land. He argues that open-field communities dealt with this risk by scattering landholdings into thin strips, making it hard for shepherds to focus on any particular plot, and that this redivision of boundaries was preferable to closer monitoring of shepherds' activities.

Online semicommons



This framework is useful for describing online communities. It leads us to ask which aspects of the resource are private, which are common, what forms of strategic behavior this combination faces, and what institutions respond to those threats. The physical infrastructure on which a community operates is rival and privately held; we're interested in those communities where that infrastructure is thrown open at the content layer, and thus effectively held in common.



Authors

Moderators

Readers

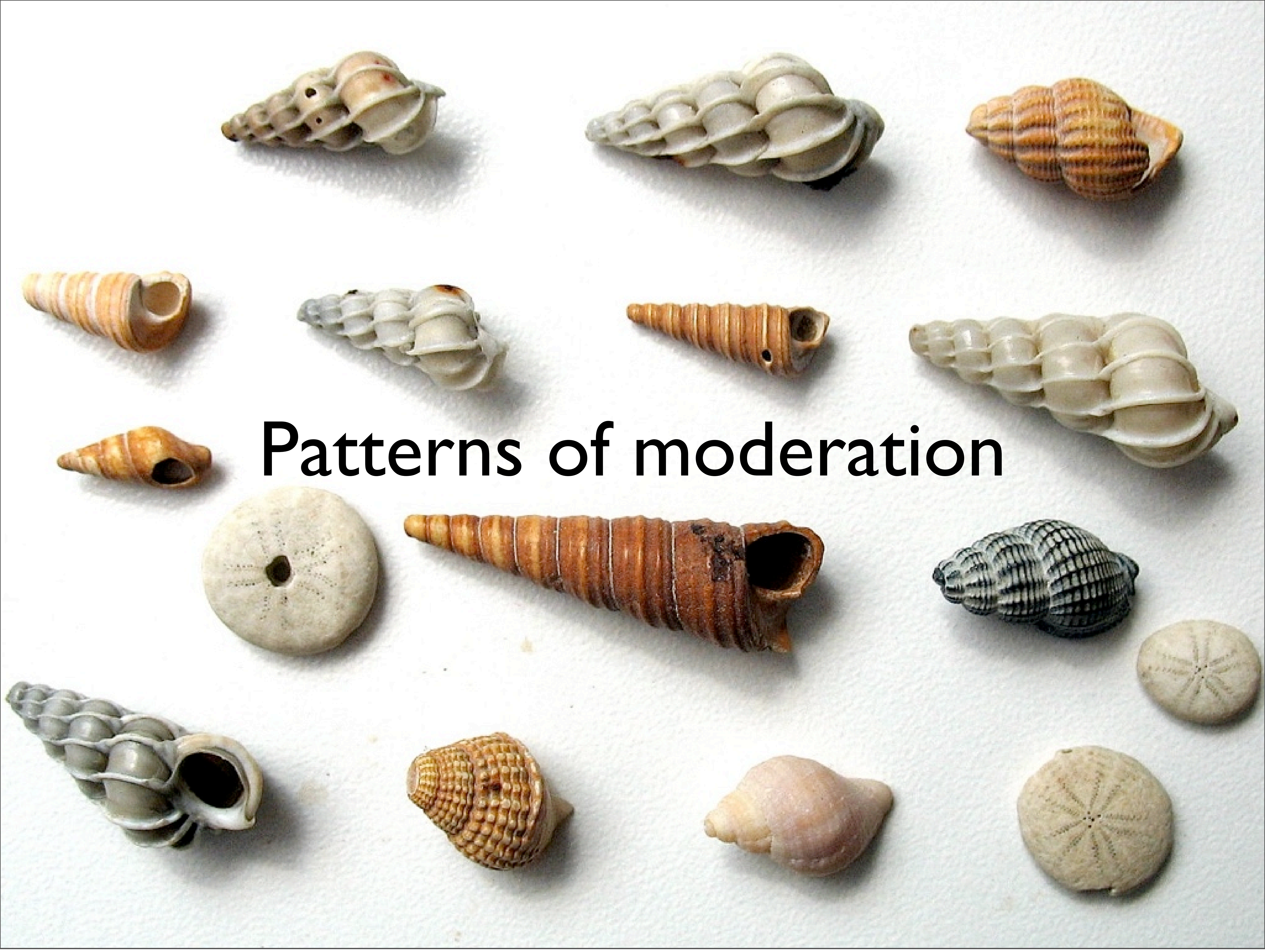
Content

Infrastructure

Owners

To think about the relevant kinds of strategic behavior, it’s useful to have a more formal model of how people use such a semicommons. I simplify by thinking of “roles”: authors produce content that readers consume. Moderators intervene to edit that content. All of these activities require support from the infrastructure, which is provided and maintained by owners. Of course, users may play more than one role.

These relationships generate several broad classes of strategic behavior. **Congestion**, primarily created by authors and to a lesser extent by readers, involves overuse of the rival infrastructure. **Cacophony** involves overuse of the nonrival content layer, making it harder for readers to find what they want. **Manipulation** involves moderators taking unfair control of what readers see; and **weaponization** involves the use of the system to produce harmful content.



Patterns of moderation

Dealing with these problems is the job of **moderation** in the broader sense. These are the techniques that a community uses to prevent abuses and create useful order. The work of “moderators” in editing content is one way that a community can do this, but the term “moderation” as I’m using it also encompasses all the other institutions of self-governance that shape how users interact.

The idea of “patterns” of moderation comes from architectural theory, where Christopher Alexander has developed the idea of a “pattern” as a recurring solution to a common design problem. The idea emphasizes that the pattern’s details change with the details of the particular instantiation of the problem, and also the way in which multiple patterns fit together in a “pattern language.” Thus, for example, some pattern of excludability of particular identified users provides a useful foundation for a pattern of pricing access. If you can’t keep out those who don’t pay, pricing alone won’t do much. My goal in the paper is to provide a useful vocabulary for talking about moderation patterns and predicting how they will work and when they will not.

Social Norms

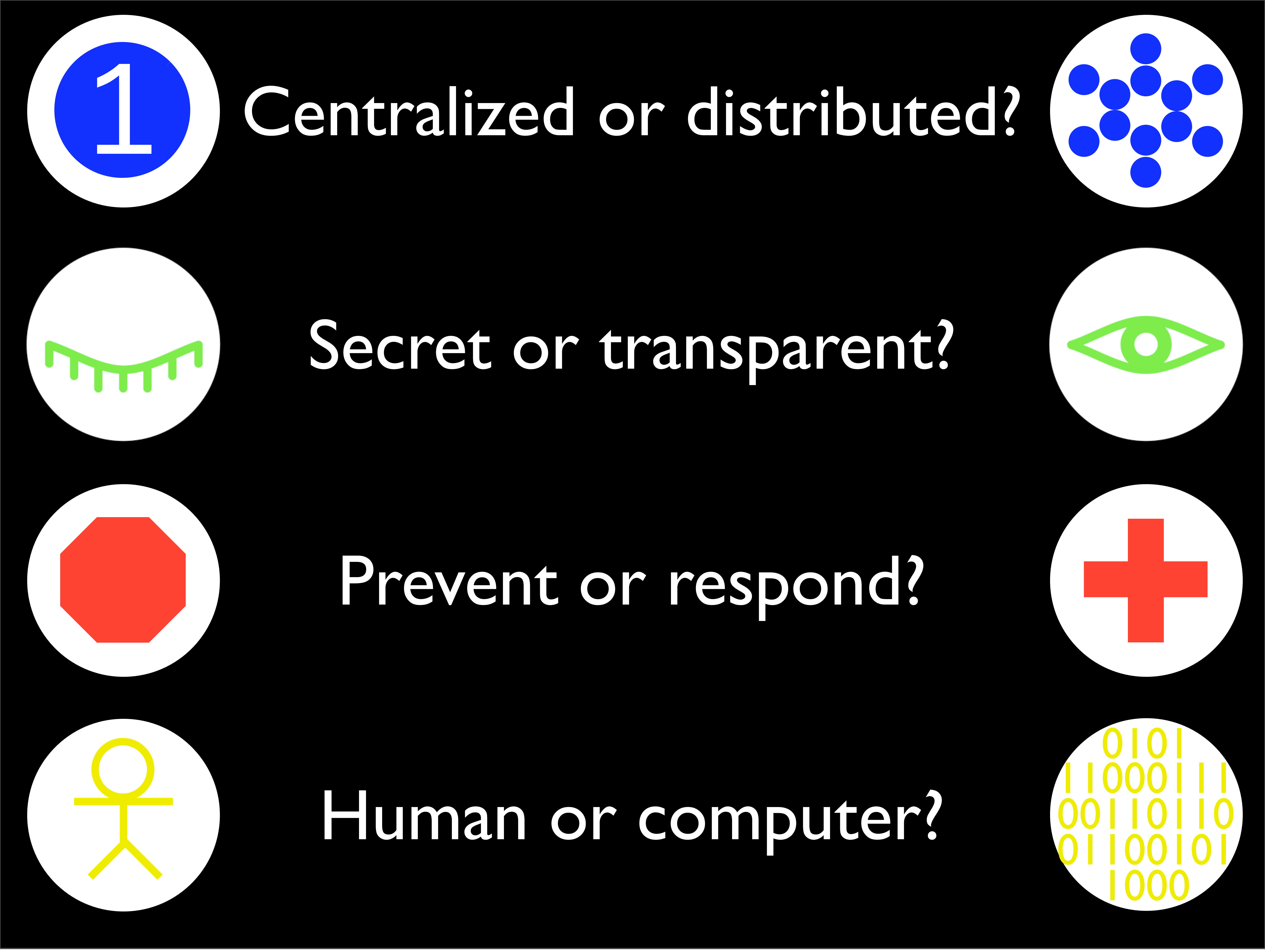
Pricing

Organization

Exclusion

I go into the elements of the patterns of moderation in some detail in the paper, including their interactions with each other. I'll discuss here only some of the moving parts, to give a feel for how one can use them in analyzing communities.

Everything starts with the four main verbs of moderation. Social norms are fundamental; a community with strong norms of collaboration is far more likely to work. They're also hard to affect directly, so almost every other moderation decision is important for its indirect effects on social norms. Pricing maps fairly directly onto Lessig's "market" modality of regulation. Exclusion involves keeping out unwanted users. Organization, the most technically complex, involves sorting, annotating, editing, and combining authors' contributions to make them more useful to readers.



Centralized or distributed?

Secret or transparent?

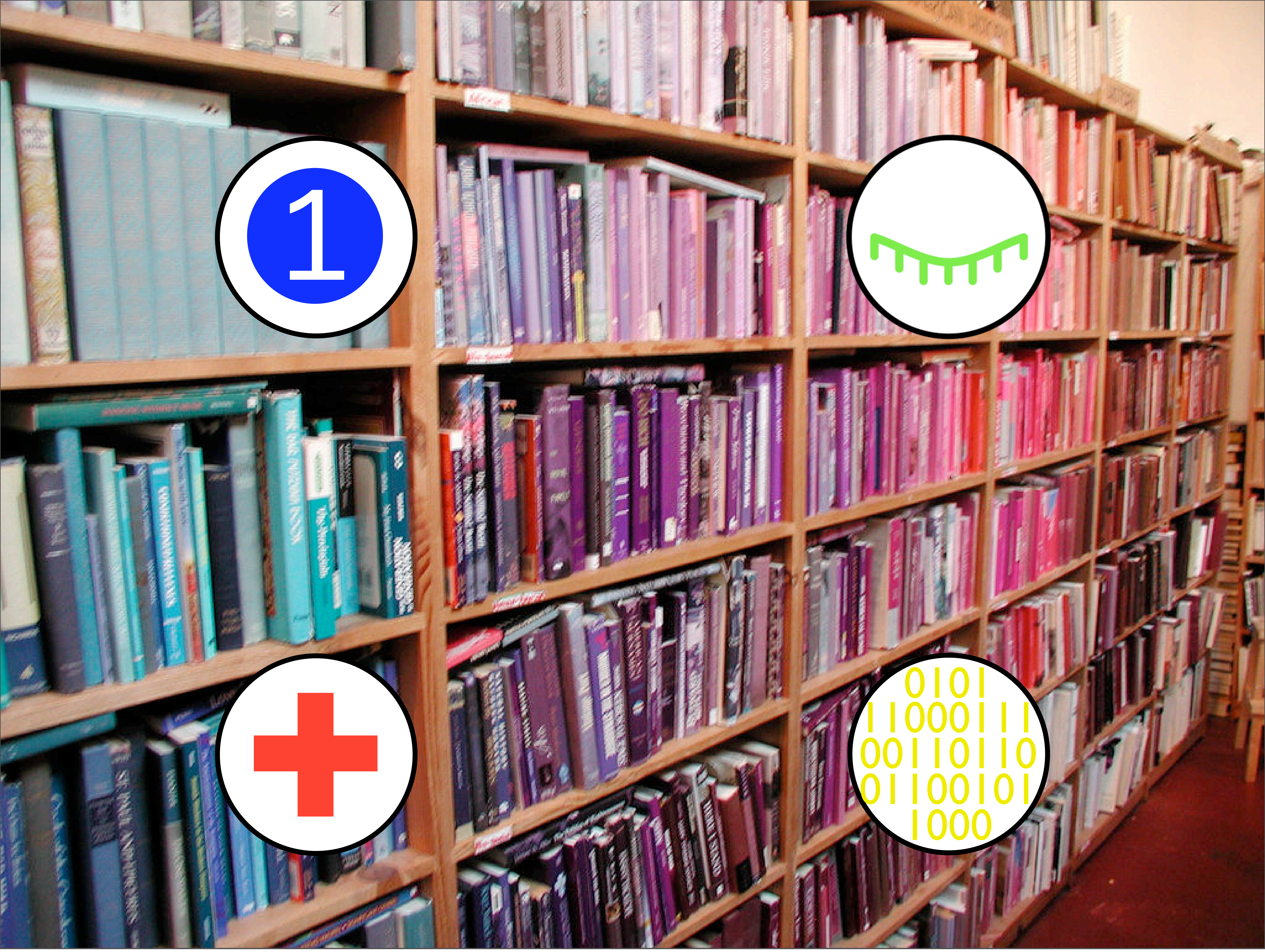
Prevent or respond?

Human or computer?

These four verbs can be implemented in an infinite variety of ways. Here are some questions that are helpful to ask in understanding a given pattern. First, is the moderation decision centralized and unitary, so that everyone sees the same results -- or is it multiple and distributed, so that different parts of the community may make different moderation decisions? Centralization provides coherence and order; distribution allows people of differing views to live and let live. Second, is the decision made for secret reasons or transparent ones? Secrecy can avoid loopholing and gamesmanship; transparency promotes legitimacy and trust. Third, is the decision an ex ante prevention of trouble or an ex post cleanup of it? Prevention can prevent unwanted content from becoming entrenched, response is often easier to implement. And fourth, are the decisions made by a person or by a computer? People can use greater discretion; computers can act more consistently and cheaply. These four characteristics can determine a great deal about how moderation works in practice.

Let’s work some brief examples.

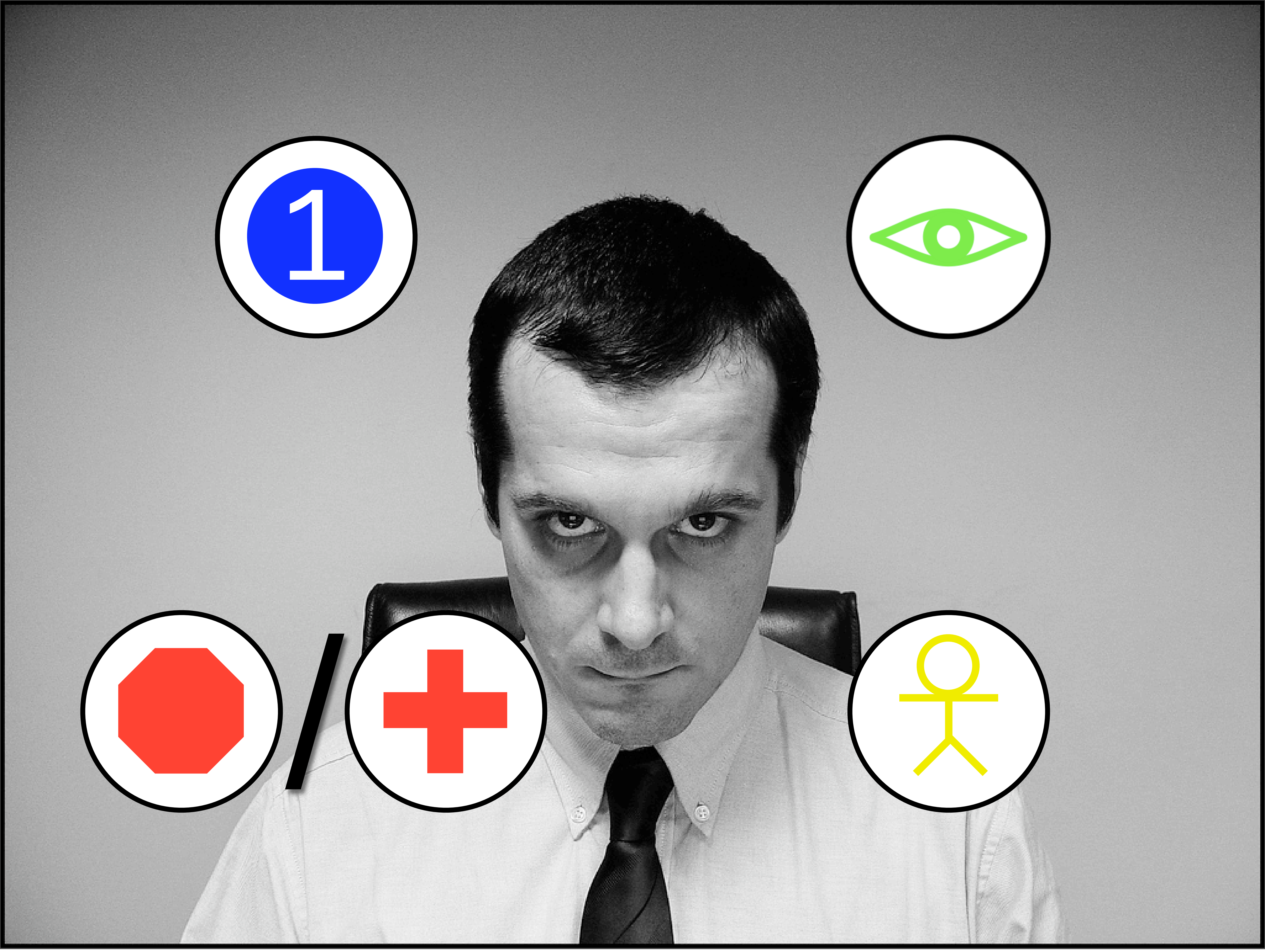
Search engines, for starters, are tools for bringing order to the web. They reduce cacophony by helping readers find only the content they’re looking for. How would we characterize search engines as a form of moderation?



I hope you’re with me when I say that search is a form of organization. Sorting content into displayed and hidden is fundamentally a form of organization. More precisely, any given search engine is centralized (since it uses a single, concentrated ranking algorithm). It’s generally secret, since the algorithm is kept private. It involves response, since the search engine finds and indexes already-existing content, rather than categorizing content before it is available on the web. And search is famously computation-intensive; huge farms of computers carry it out.

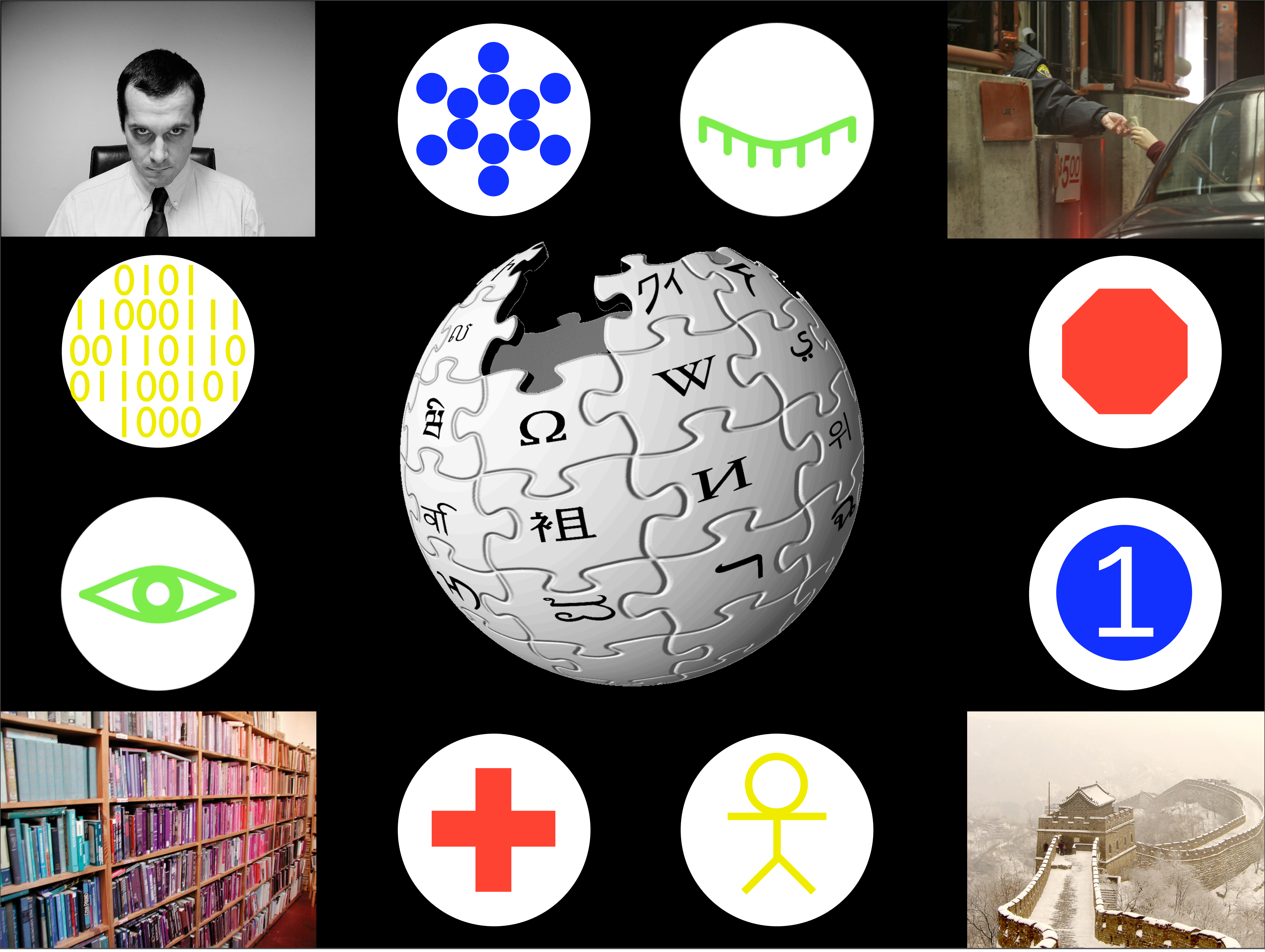
artmusicflash
politics photography
science history bush humor
Religion technology election law video war
game games film fun weird books culture funny movies
lesign food photos iraq language amusing advertising space
omics architecture gallery animation tv travel silly internet UK media
ews blogs terrorism gay canada apple 9-11 MP3 literature television drugs
log sex death marketing health linguistics copyright conspiracy toys utah illustration japan
iology movie children education NSFW radio America economics philosophy sports wacky christmas
artoons deadlink Government journalism comedy words google women israel collections NewYork researc
urope showcase humour poetry US rights russia audio collection environment satire bbc retro parody mathematics
teractive security physics animals magazine alcohol africa puzzles islam france crime propaganda web stupid writing
ction privacy webcomics maps museum microsoft evolution mystery fraud starwars blogging china censorship jazz distraction
omputers celebrity urban AIDS Asia images software fashion WWII ebay gender intellectualproperty library photo everydayobjects
?
METAFILTER
COMPILATION
ALBUM (VOL. Ø1)
IS AT THE DUPLICATORS
ETA NOVEMBER 2006
READ
RESPONSIBL

By way of contrast, consider Metafilter. That’s a community weblog whose users post links and discussion. It’s run, more or less, by three people. How do they do it?

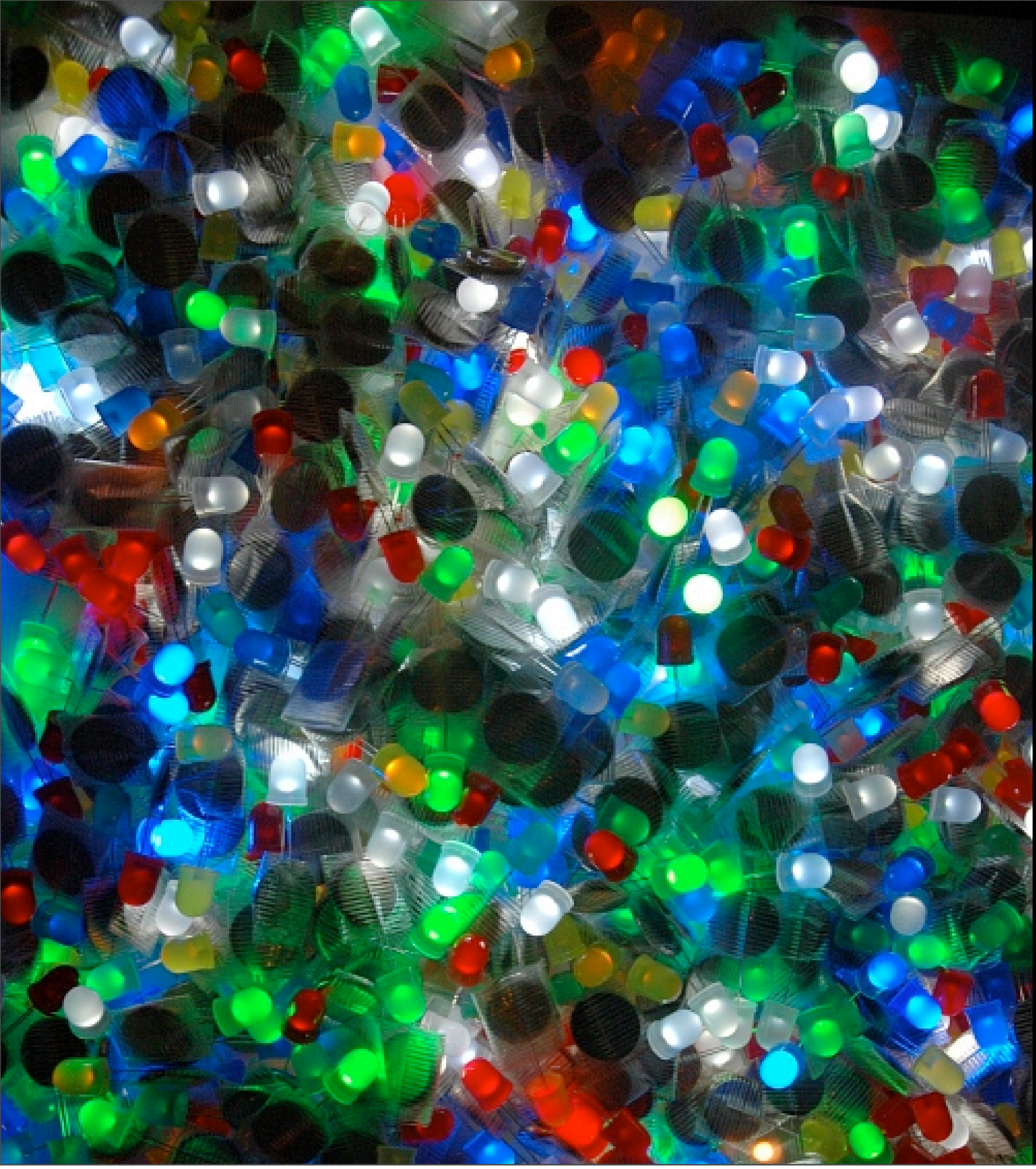


Now, as I describe in the paper, they use a lot of different techniques as appropriate. But if you ask Matt Haughey, the founder and main administrator, what he's doing, he'll tell you that it's about about creating positive social norms (although he'll use a more descriptive term, like "community spirit.") He and Jessamyn and Josh spend most of their time communicating with other users to create a feeling of cooperation and respectfulness. That's centralized, because they're administrators for the whole site trying to create a unified feel. It's transparent; they explain themselves at great length. It's both preventive and responsive; they deal with outbreaks and try to set better norms for the future. And it's deeply deeply human.

What about Wikipedia, that charnel house of knowledge?



If you look closely, you can find examples of everything I’ve described. IP banning of large-scale vandals involves centralized exclusion; comments on someone’s talk page are distributed norm-setting; there are vigilante human editors and officious bots; there are retroactive cleanups and a million policies going forward; some things are argued to death and others are mysterious . . . it’s all in there. Indeed, I think part of the power of Wikipedia is that its community of users is seething with ideas and conversation, making it profoundly generative in developing new responses to threats.



Moderation comes in a wide variety of forms.

So let’s take stock of the main lessons from adopting a pattern-oriented view of moderation in online communities.

First, there is the basic point that moderation is not one phenomenon, but rather a whole phylum of techniques. Different forms of moderation have very different properties and they work together and against each other in complicated ways. This diversity should make us reluctant to generalize about what all forms of moderation do or do not do. Much depends on the particulars of the situation and on the particular forms in use.

Choosing among them requires trading off virtues.

Each pattern has advantages and disadvantages. Centralized computer ex ante secret exclusion can be highly unfair to those excluded and can keep group size under the critical mass for large-scale creation. On the other hand, it can be very easy to enforce and can help promote cooperative norms among those who are included. Sometimes this tradeoff is a good one; sometimes it is not; often the choice can be fairly debated. The point is that moderation intrinsically involves sacrificing some facets of openness in order to preserve and enhance others.

Individual patterns are fragile.

Some forms of moderation, such as norms of low posting volume, break down under determined attack. Some, such as human ex ante organization, do not scale well. Some, such as centralized human pricing, can face serious information problems. No individual pattern of moderation will always work. Successful communities are more often those that have dynamic ability to adapt their patterns to their challenges than those that picked the absolute right pattern for all time.

Sometimes moderation fails.

Some communities make bad choices, or fail to adapt when their old patterns are failing. Usenet worked for a long time, but its particular patterns of distributed organization proved susceptible to spam and to malicious cross-posting. The Internet is littered with Slashdot clones -- many using the same software powering Slashdot -- that crashed and burned. There are blogs and mailing lists whose moderators have alienated most of the regular contributors. There are many ways to fail.

Sometimes moderation works.

But there are also many ways to succeed. This is, after all, the story of the Internet. It is filled with sites whose communities of users are healthy and thriving, that manage to be substantially open. The email system more or less works. Wikipedia is growing by leaps and bounds. My weblog has open comments; I don't have any significant problems. I subscribe to a dozen mailing lists that all pretty much take care of themselves. The web as a whole is still a great place of astonishing ferment. At its best, moderation can produce sustainable, healthy communities. What I've tried to do is help make some progress in understanding how.

Discussion

Colophon

- The following photographs were used by permission, are © their respective photographers and, where noted, are available under the indicated Creative Commons license. All other elements are © James Grimmelmann and are available under a Creative Commons Attribution 3.0 United States license.

- Paving stones: 1Sock, “Break It Down For Me,” http://www.flickr.com/photos/1sock/301276274/, BY-NC-ND 2.0
- Ram: John Baird, “Ram,” http://www.flickr.com/photos/johnbaird/145718528/, BY-NC-ND 2.0
- Photomosaic: Jim Bumgardner, “Mosaic Portrait: Other Things, ”http://www.flickr.com/photos/krazydad/74287919/, BY-NC-SA 2.0
- Seashells: Supermietzi, “1980,” http://www.flickr.com/photos/supermietzi/188063021/, BY-NC-SA 2.0
- Writing hand: kirstenv, http://www.flickr.com/photos/kirstenverstraten/443978021/, BY-NC-ND 2.0
- Reading: Gwen Harlow, “D'arby reading Beloved,” http://www.flickr.com/photos/gwen/357714932/, BY-NC-ND 2.0
- Lineman: David Nelson, “Linemen_2,” http://www.flickr.com/photos/dleroy/6704942/, BY-NC-SA 2.0
- Hockey fight: DarbCU, “Fight,” http://www.flickr.com/photos/darbcu/262744282/, BY-NC-ND 2.0
- Great wall of China: Steve Webel, “Great Wall of China (2004),” http://www.flickr.com/photos/webel/64438599/, BY-ND 2.0
- Tollbooth: Jordansmall, “The Handoff,” http://www.flickr.com/photos/jordansmall/88129701/, BY-NC-ND 2.0
- Frowning man: schizoo23, “the main man,” http://www.flickr.com/photos/schizoo23/360887176/, BY-NC-SA 2.0
- Books: Mermaniac, “Rainbow of Books 3,” http://www.flickr.com/photos/mermaniac/1473992/, BY-NC-ND 2.0
- Jessamyn West: redjar, “Jessamyn West,” http://www.flickr.com/photos/redjar/113978938/ BY-SA 2.0
- Matt Haughey: Matt Haughey, “Posted at 19.6 mph,” http://www.flickr.com/photos/mathowie/10774597/, BY-NC-SA 2.0
- Josh Mllard: Kathleen Bennett, “Cortex is really, really, working on the compilation album,’ http://flickr.com/photos/ferneyes/283066436/, BY-NC-ND 2.0
- Scales: Esther Dyson, “weight 2.0,” http://www.flickr.com/photos/edyson/87566058/, BY-NC 2.0
- LEDS: disposable fragments, “LED throwies,” http://www.flickr.com/photos/0816style/235198839/, BY-NC-SA 2.0
- Droplets: Steve Wall, “fragile universe,” http://www.flickr.com/photos/stevewall/188024983/, BY-NC-SA 2.0
- Interior Ruins: John Baird, “Fisher Body Plant 21,” http://www.flickr.com/photos/johnbaird/112856502/, BY-NC-ND 2.0
- Ping’an Village: Kevin Lam, “Ping’an Village,” http://www.flickr.com/photos/lamkevin/272844361/, all rights reserved