5

Tethered Appliances, Software as Service, and Perfect Enforcement

As Part I of this book explained, the generative nature of the PC and Internet—a certain incompleteness in design, and corresponding openness to outside innovation—is both the cause of their success and the instrument of their forthcoming failure.

The most likely reactions to PC and Internet failures brought on by the proliferation of bad code, if they are not forestalled, will be at least as unfortunate as the problems themselves. People now have the opportunity to respond to these problems by moving away from the PC and toward more centrally controlled—"tethered"—information appliances like mobile phones, video game consoles, TiVos, iPods, iPhones, and BlackBerries. The ongoing communication between this new generation of devices and their vendors assures users that functionality and security improvements can be made as new problems are found. To further facilitate glitch-free operation, devices are built to allow no one but the vendor to change them. Users are also now able to ask for the appliancization of their own PCs, in the process forfeiting the ability to easily install new code themselves. In a development reminiscent of the old days of AOL and CompuServe, it

is increasingly possible to use a PC as a mere dumb terminal to access Web sites with interactivity but with little room for tinkering. ("Web 2.0" is a new buzzword that celebrates this migration of applications traditionally found on the PC onto the Internet. Confusingly, the term also refers to the separate phenomenon of increased user-generated content and indices on the Web—such as relying on user-provided tags to label photographs.) New information appliances that are tethered to their makers, including PCs and Web sites refashioned in this mold, are tempting solutions for frustrated consumers and businesses.

None of these solutions, standing alone, is bad, but the aggregate loss will be enormous if their emergence represents a wholesale shift of our information ecosystem away from generativity. Some are skeptical that a shift so large can take place.[1] But confidence in the generative Internet's inertia is misplaced. It discounts the power of fear should the existing system falter under the force of particularly well-written malware. People might argue about the merits of one platform compared to another ("Linux never needs to be rebooted"),[2] but the fact is that no operating system is perfect, and, more importantly, any PC open to running third-party code at the user's behest can fail when poor code is adopted. The fundamental problem arises from too much functionality in the hands of users who may not exercise it wisely: even the safest Volvo can be driven into a wall.

People are frustrated by PC kinks and the erratic behavior they produce. Such unexpected variations in performance have long been smoothed out in refrigerators, televisions, mobile phones, and automobiles. As for PCs, telling users that their own surfing or program installation choices are to blame understandably makes them no less frustrated, even if they realize that a more reliable system would inevitably be less functional—a trade-off seemingly not required by refrigerator improvements. Worse, the increasing reliance on the PC and Internet that suggests momentum in their use means that more is at risk when something goes wrong. Skype users who have abandoned their old-fashioned telephone lines may regret their decision if an emergency arises and they need to dial an emergency number like 911, only to find that they cannot get through, let alone be located automatically.[3] When one's finances, contacts, and appointments are managed using a PC, it is no longer merely frustrating if the computer comes down with a virus. It is enough to search for alternative architectures.

A shift to tethered appliances and locked-down PCs will have a ripple effect on long-standing cyberlaw problems, many of which are tugs-of-war between individuals with a real or perceived injury from online activity and those who

wish to operate as freely as possible in cyberspace. The capacity for the types of disruptive innovation discussed in the previous chapter will not be the only casualty. A shift to tethered appliances also entails a sea change in the *regulability* of the Internet. With tethered appliances, the dangers of excess come not from rogue third-party code, but from the much more predictable interventions by regulators into the devices themselves, and in turn into the ways that people can use the appliances.

The most obvious evolution of the computer and network—toward tethered appliancization—is on balance a bad one. It invites regulatory intervention that disrupts a wise equilibrium that depends upon regulators acting with a light touch, as they traditionally have done within liberal societies.

THE LONG ARM OF MARSHALL, TEXAS

TiVo introduced the first digital video recorder (DVR) in 1998.[4] It allowed consumers to record and time-shift TV shows. After withstanding several claims that the TiVo DVR infringed other companies' patents because it offered its users on-screen programming guides,[5] the hunted became the hunter. In 2004, TiVo sued satellite TV distributor EchoStar for infringing TiVo's own patents[6] by building DVR functionality into some of EchoStar's dish systems.[7]

A Texas jury found for TiVo. TiVo was awarded $90 million in damages and interest. In briefs filed under seal, TiVo apparently asked for more. In August 2006, the court issued the following ruling:

> Defendants are hereby . . . ORDERED to, within thirty (30) days of the issuance of this order, disable the DVR functionality (i.e., disable all storage to and playback from a hard disk drive of television data) in all but 192,708 units of the Infringing Products that have been placed with an end user or subscriber.[8]

That is, the court ordered EchoStar to kill the DVR functionality in products already owned by "end users": millions of boxes which were already sitting in living rooms around the world[9] with owners who might be using them at that very instant.[10] Imagine sitting down to watch television on an EchoStar box, and instead finding that all your recorded shows had been zapped, along with the DVR functionality itself—killed by remote signal traceable to the stroke of a judge's quill in Marshall, Texas.

The judicial logic for such an order is drawn from fundamental contraband rules: under certain circumstances, if an article infringes on intellectual prop-

erty rights, it can be impounded and destroyed.[11] Impoundment remedies are usually encountered only in the form of Prohibition-era-style raids on warehouses and distribution centers, which seize large amounts of contraband before it is sold to consumers.[12] There are no house-to-house raids to, say, seize bootleg concert recordings or reclaim knockoff Rolexes and Louis Vuitton handbags from the people who purchased the goods.

TiVo saw a new opportunity in its patent case, recognizing that EchoStar's dish system is one of an increasing number of modern tethered appliances. The system periodically phones home to EchoStar, asking for updated programming for its internal software.[13] This tethered functionality also means EchoStar can remotely destroy the units. To do so requires EchoStar only to load its central server with an update that kills EchoStar DVRs when they check in for new features.

As of this writing, *TiVo v. EchoStar* is pending appeal on other grounds.[14] The order has been stayed, and no DVRs have yet been remotely destroyed.[15] But such remote remedies are not wholly unprecedented. In 2001, a U.S. federal court heard a claim from a company called PlayMedia that AOL had included PlayMedia's AMP MP3 playback software in version 6.0 of AOL's software in violation of a settlement agreement between PlayMedia and a company that AOL had acquired. The court agreed with PlayMedia and ordered AOL to prevent "any user of the AOL service from completing an online 'session' . . . without AMP being removed from the user's copy of AOL 6.0 by means of an AOL online 'live update.'"[16]

TiVo v. EchoStar and *PlayMedia v. AOL* broach the strange and troubling issues that arise from the curious technological hybrids that increasingly populate the digital world. These hybrids mate the simplicity and reliability of television-like appliances with the privileged power of the vendor to reprogram those appliances over a network.

REGULABILITY AND THE TETHERED APPLIANCE

As legal systems experienced the first wave of suits arising from use of the Internet, scholars such as Lawrence Lessig and Joel Reidenberg emphasized that code could be law.[17] In this view, the software we use shapes and channels our online behavior as surely as—or even more surely and subtly than—law itself. Restrictions can be enforced by the way a piece of software operates. Our ways of thinking about such "west coast code"[18] are still maturing, and our instincts

for when we object to such code are not well formed. Just as technology's functionality defines the universe in which people can operate, it also defines the range of regulatory options reasonably available to a sovereign. A change in technology can change the power dynamic between those who promulgate the law and those who are subject to it.[19]

If regulators can induce certain alterations in the nature of Internet technologies that others could not undo or widely circumvent, then many of the regulatory limitations occasioned by the Internet would evaporate. Lessig and others have worried greatly about such potential changes, fearing that blunderbuss technology regulation by overeager regulators will intrude on the creative freedom of technology makers and the civic freedoms of those who use the technology.[20]

So far Lessig's worries have not come to pass. A system's level of generativity can change the direction of the power flow between sovereign and subject in favor of the subject, and generative Internet technology has not been easy to alter. There have been private attempts to use code to build so-called trusted systems, software that outsiders can trust to limit users' behavior—for example, by allowing a song to be played only three times before it "expires," or by preventing an e-book from being printed.[21] (Code-based enforcement mechanisms are also variously called digital rights management systems or technological protection measures.)[22] Most trusted systems have failed, often because either savvy users have cracked them early on or the market has simply rejected them. The few that have achieved some measure of adoption—like Apple iTunes's FairPlay, which allows purchased songs to exist on only five registered devices at once[23]—are either readily circumvented, or tailored so they do not prevent most users' desired behavior.

Even the governments most determined to regulate certain flows of information—such as China—have found it difficult to suppress the flow of data on the Internet.[24] To be sure, with enough effort, censorship can have some effect, especially because most citizens prefer to slow down for speed bumps rather than invent ways around them.[25] When a Web site fails to load, for example, users generally visit a substitute site rather than wait. Taking advantage of this reality, Chinese regulators have used their extensive control over ISPs' routing of data packets to steer users away from undesirable Web sites by simply causing the Web pages to fail to load in the course of normal surfing.

But so long as the *endpoints* remain generative and any sort of basic Internet access remains available, subversively minded techies can make applications

that offer a way around network blocks.[26] Such applications can be distributed through the network, and unsavvy users can then partake simply by double-clicking on an icon. Comprehensive regulatory crackdowns require a non-generative endpoint or influence over the individual using it to ensure that the endpoint is not repurposed.

For example, non-generative endpoints like radios and telephones can be constrained by filtering the networks they use. Even if someone is unafraid to turn a radio tuning knob or dial a telephone number to the outside world, radio broadcasts can be jammed, and phone connections can be disabled or monitored. Because radios and telephones are not generative, such jamming cannot be circumvented. North Korea has gone even further with endpoint lockdown. There, by law, the radios themselves are built so that they cannot be tuned to frequencies other than those with official broadcasts.[27]

With generative devices like PCs, the regulator must settle for either much leakier enforcement or much more resource-intensive measures that target the individual—such as compelling citizens to perform their Internet surfing in cyber cafés or public libraries, where they might limit their activities for fear that others are watching.

The shift toward non-generative endpoint technology driven by consumer security worries of the sort described in this book changes the equation.[28] The traditional appliance, or nearly any object, for that matter, once placed with an individual, belongs to that person. Tethered appliances belong to a new class of technology. They are *appliances* in that they are easy to use, while not easy to tinker with. They are *tethered* because it is easy for their vendors to change them from afar, long after the devices have left warehouses and showrooms. Consider how useful it was in 2003 that Apple could introduce the iTunes Store directly into iTunes software found on PCs running Mac OS.[29] Similarly, consumers can turn on a TiVo—or EchoStar—box to find that, thanks to a remote update, it can do new things, such as share programs with other televisions in the house.[30]

These tethered appliances receive remote updates from the manufacturer, but they generally are not configured to allow anyone else to tinker with them—to invent new features and distribute them to other owners who would not know how to program the boxes themselves. Updates come from only one source, with a model of product development limited to non-user innovation. Indeed, recall that some recent devices, like the iPhone, are updated in ways that actively seek out and erase any user modifications. These boxes thus resemble the early proprietary information services like CompuServe and AOL,

for which only the service providers could add new features. Any user inventiveness was cabined by delays in chartering and understanding consumer focus groups, the hassles of forging deals with partners to invent and implement suggested features, and the burdens of performing technical R&D.

Yet tethered appliances are much more powerful than traditional appliances. Under the old regime, a toaster, once purchased, remains a toaster. An upgraded model might offer a third slot, but no manufacturer's representative visits consumers and retrofits old toasters. Buy a record and it can be played as many times as the owner wants. If the original musician wishes to rerecord a certain track, she will have to feature it in a successive release—the older work has been released to the four winds and cannot be recalled.[31] A shift to smarter appliances, ones that can be updated by—and only by—their makers, is fundamentally changing the way in which we experience our technologies. Appliances become *contingent:* rented instead of owned, even if one pays up front for them, since they are subject to instantaneous revision.

A continuing connection to a producer paves the way for easier postacquisition improvements: the modern equivalent of third slots for old toasters. That sounds good: more features, instantly distributed. So what is the drawback? Those who believe that markets reflect demand will rightly ask why a producer would make post hoc changes to technology that customers may not want.

One answer is that they may be compelled to do so. Consider EchoStar's losing verdict in Marshall, Texas. If producers can alter their products long after the products have been bought and installed in homes and offices, it occasions a sea change in the *regulability* of those products and their users. With products tethered to the network, regulators—perhaps on their own initiative to advance broadly defined public policy, or perhaps acting on behalf of parties like TiVo claiming private harms—finally have a toolkit for exercising meaningful control over the famously anarchic Internet.

TYPES OF PERFECT ENFORCEMENT

The law as we have known it has had flexible borders. This flexibility derives from prosecutorial and police discretion and from the artifice of the outlaw. When code is law, however, execution is exquisite, and law can be self-enforcing. The flexibility recedes. Those who control the tethered appliance can control the behavior undertaken with the device in a number of ways: preemption, specific injunction, and surveillance.

Preemption

Preemption entails anticipating and designing against undesirable conduct before it happens. Many of the examples of code as law (or, more generally, architecture as law) fit into this category. Lessig points out that speeding can be regulated quite effectively through the previously mentioned use of speed bumps.[32] Put a speed bump in the road and people slow down rather than risk damaging their cars. Likewise, most DVD players have Macrovision copy protection that causes a signal to be embedded in the playback of DVDs, stymieing most attempts to record DVDs onto a VCR.[33] Owners of Microsoft's Zune music player can beam music to other Zune owners, but music so transferred can be played only three times or within three days of the transfer.[34] This kind of limitation arguably preempts much of the damage that might otherwise be thought to arise if music subject to copyright could be shared freely. With TiVo, a broadcaster can flag a program as "premium" and assign it an expiration date.[35] A little red flag then appears next to it in the viewer's list of recorded programs, and the TiVo will refuse to play the program after its expiration date. The box's makers (or regulators of the makers) could further decide to automatically reprogram the TiVo to limit its fast-forwarding functionality or to restrict its hours of operability. (In China, makers of multiplayer games have been compelled to limit the number of hours a day that subscribers can play in an effort to curb gaming addiction.)[36] Preemption does not require constant updates so long as the device cannot easily be modified once it is in the user's possession; the idea is to design the product with broadly defined limits that do not require further intervention to serve the regulator's or designer's purposes.

Specific Injunction

Specific injunction takes advantage of the communication that routinely occurs between a particular tethered appliance and its manufacturer, *after* it is in consumer hands, to reflect changed circumstances. The *TiVo v. EchoStar* remedy belongs in this category, as it mandates modification of the EchoStar units after they have already been designed and distributed. This remote remedy was practicable because the tethering allowed the devices to be completely reprogrammed, even though the initial design of the EchoStar device had not anticipated a patent infringement judgment.

Specific injunction also allows for much more tailored remedies, like the PlayMedia-specific court order discussed earlier. Such tailoring can be content-

specific, user-specific, or even time-specific. These remedies can apply to some units and not others, allowing regulators to winnow out bad uses from good ones on the basis of individual adjudication, rather than rely on the generalities of *ex ante* legislative-style drafting. For example, suppose a particular television broadcast were found to infringe a copyright or to damage someone's reputation. In a world of old-fashioned televisions and VCRs, or PCs and peer-to-peer networks, the broadcaster or creator could be sued, but anyone who recorded the broadcast could, as a practical matter, retain a copy. Today, it is possible to require DVR makers to delete the offending broadcast from any DVRs that have recorded it or, perhaps acting with more precision, to retroactively edit out the slice of defamatory content from the recorded program. This control extends beyond any particular content medium: as e-book devices become popular, the same excisions could be performed for print materials. Tailoring also could be user-specific, requiring, say, the prevention or elimination of prurient material from the devices of registered sex offenders but not from others' devices.

Surveillance

Tethered appliances have the capacity to relay information about their uses back to the manufacturer. We have become accustomed to the idea that Web sites track our behavior when we access them—an online bookseller, for example, knows what books we have browsed and bought at its site. Tethered appliances take this knowledge a step further, recording what we do with the appliances even in transactions that have nothing to do with the vendor. A TiVo knows whether its owner watches FOX News or PBS. It knows when someone replays some scenes and skips others. This information is routinely sent to the TiVo mothership;[37] for example, in the case of Janet Jackson's "wardrobe malfunction" during the 2004 Super Bowl halftime show, TiVo was able to calculate that this moment was replayed three times more frequently than any other during the broadcast.[38]

TiVo promises not to release such surveillance information in personally identifiable form, but the company tempers the promise with an industry-standard exception for regulators who request it through legal process.[39] Automakers General Motors and BMW offer similar privacy policies for the computer systems, such as OnStar, built into their automobiles. OnStar's uses range from providing turn-by-turn driving directions with the aid of Global Positioning System (GPS) satellites, to monitoring tire pressure, providing emergency assistance, and facilitating hands-free calling with embedded microphones and

speakers. The FBI realized that it could eavesdrop on conversations occurring inside an OnStar-equipped vehicle by remotely reprogramming the system to activate its microphones for use as a "roving bug," and it has secretly ordered an anonymous carmaker to do just that on at least one occasion.[40]

A similar dynamic is possible with nearly all mobile phones. Mobile phones can be reprogrammed at a distance, allowing their microphones to be secretly turned on even when the phone is powered down. All ambient noise and conversation can then be continuously picked up and relayed back to law enforcement authorities, regardless of whether the phone is being used for a call.[41] On modern PCs equipped with an automatic update feature, there is no technical barrier that prevents the implementation of any similar form of surveillance on the machine, whether it involves turning on the PC's microphone and video camera, or searching and sharing any documents stored on the machine. Such surveillance could be introduced through a targeted update from the OS maker or from any other provider of software running on the machine.

Surveillance need not be limited to targeted eavesdropping that is part of a criminal or civil investigation. It can also be effected more generally. In 1996, law student Michael Adler offered the hypothetical of an Internet-wide search for contraband.[42] He pointed out that some digital items might be illegal to possess or be indicative of other illegal activity—for example, child pornography, leaked classified documents, or stores of material copied without permission of the copyright holder. A Net-wide search could be instigated that would inventory connected machines and report back when smoking guns were found.

Tethering makes these approaches practicable and inexpensive for regulators. A government need only regulate certain critical private intermediaries—those who control the tethered appliances—to change the way individuals experience the world. When a doctrine's scope has been limited by prudential enforcement costs, its reach can be increased as the costs diminish.

EVALUATING PERFECT ENFORCEMENT

The prospect of more thorough or "perfect" law enforcement may seem appealing. If one could wave a wand and make it impossible for people to kill each other, there might seem little reason to hesitate. Although the common law has only rarely sought to outright prohibit the continued distribution of defamatory materials by booksellers and newsstands, much less continued possession

by purchasers, ease of enforcement through tethered appliances could make it so that all such material—wherever it might be found—could vanish into the memory hole. Even when it comes to waving the regulator's wand for the purpose of eradicating online evils like harassment, invasion of privacy, and copyright infringement, there are important reasons to hesitate.[43]

Objections to the Underlying Substantive Law

Some people are consistently diffident about the presence of law in the online space. Those with undiluted libertarian values might oppose easier enforcement of laws as a general matter, because they believe that self-defense is the best solution to harm by others, especially within a medium that carries bits, not bullets.[44] By these lights, the most common online harms simply are not as harmful as those in the physical world, and therefore they call for lesser intrusions. For example, defamatory speech might be met not by a lawsuit for money damages or an injunction requiring deletion of the lies, but rather by more speech that corrects the record. A well-configured e-mail client can adequately block spam, making it unnecessary to resort to intervention by a public authority. Material harmful to minors can be defanged by using parental filters, or by providing better education to children about what to expect when they go online and how to deal with images of violence and hate.

Such "just deal with it" arguments are deployed less often against the online circulation of images of child abuse. The creation and distribution of child pornography is nearly universally understood as a significant harm. In this context, those arguing in favor of an anarchic environment shift to claims that the activity is not very common or that existing tools and remedies are sufficiently effective—or they rely on some of the other objections described below.

One can also argue against stronger enforcement regimes by objecting to the laws that will be enforced. For example, many of those who argue against increased copyright enforcement—undertaken through laws that broaden infringement penalties[45] or through trusted systems that preempt infringement[46]—argue that copyright law itself is too expansive.[47] For those who believe that intellectual property rights have gone too far, it is natural to argue against regimes that make such rights easier to enforce, independent of seeking to reform the copyright law itself. Similarly, those who believe in lower taxes might object to a plan that makes it easier for intermediaries to collect and remit use and sales taxes for online transactions.[48] Likewise, the large contingent of people who routinely engage in illegal online file sharing may naturally dis-

favor anything that interferes with these activities.[49] To be sure, some of those people may download even though they believe it to be wrong—in which case they might welcome a system that better prevents them from yielding to temptation.

Law professor William Stuntz notes the use of legal procedure—evolving doctrines of Fourth and Fifth Amendment protection—as a way of limiting the substantive application of unpopular laws in eighteenth- and nineteenth-century America such as those involving first heresy and sedition, and later railroad and antitrust regulation.[50] In that context, he argues, judges interpreted the Fourth and Fifth Amendments in ways designed to increase the costs to law enforcement of collecting evidence from private parties. When the judiciary began defining and enforcing a right to privacy that limited the sorts of searches police could undertake, it became more difficult to successfully prosecute objectionable crimes like heresy, sedition, or trade offenses: "It is as if privacy protection were a proxy for something else, a tool with which courts or juries could limit the government's substantive power."[51] Challenging the rise of tethered appliances helps maintain certain costs on the exercise of government power—costs that reduce the enforcement of objectionable laws.

The drawback to arguing generally against perfect enforcement because one objects to the laws likely to be enforced is that it preaches to the choir. Certainly, those who oppose copyright laws will also oppose changes to code that facilitate the law's online enforcement. To persuade those who are more favorably disposed to enforcement of substantive laws using tethered appliances, we must look to other objections.

Portability and Enforceability Without the Rule of Law

While it might be understandable that those opposed to a substantive law would also favor continued barriers to its enforcement, others might say that the price of living under the rule of law is that law ought to be respected, even if one disagrees with it. In this view, the way to protest an undesirable law is to pursue its modification or repeal, rather than to celebrate the difficulty of its enforcement.[52] The rise of procedural privacy limits described by Stuntz was itself an artifact of the law—the decisions of judges with license to interpret the Constitution. This legally sanctioned mandate is distinct from one allowing individuals to flout the law when they feel like it, simply because they cannot be easily prevented from engaging in the illicit act or caught.

But not every society operates according to a framework of laws that are democratically promulgated and then enforced by an independent judiciary.

Governments like those of China or Saudi Arabia might particularly benefit from technological configurations that allow for inexpensive surveillance or the removal of material authored by political dissidents. In a world where tethered appliances dominate, the cat-and-mouse game tilts toward the cat. Recall that the FBI can secretly eavesdrop on any automobile with an OnStar navigation system by obtaining a judge's order and ensuring that the surveillance does not otherwise disrupt the system's functioning. In a place without the rule of law, the prospect of cars rolling off the assembly line surveillance-ready is particularly unsettling. China's government has already begun experimenting with these sorts of approaches. For example, the PC telephone program Skype is not amenable to third-party changes and is tethered to Skype for its updates. Skype's distribution partner in China has agreed to censor words like "Falun Gong" and "Dalai Lama" in its text messaging for the Chinese version of the program.[53] Other services that are not generative at the technical layer have been similarly modified: Google.cn is censored by Google at the behest of the Chinese government, and Microsoft's MSN Spaces Chinese blog service automatically filters out sensitive words from blog titles.[54]

There is an ongoing debate about the degree to which firms chartered in freer societies should assist in censorship or surveillance taking place in less free societies.[55] The argument considered here is one layer deeper than that debate: if the information ecosystem at the cutting edge evolves into one that is not generative at its core, then authoritarian governments will naturally inherit an ability to enforce their wills more easily, without needing to change technologies and services or to curtail the breadth of their influence. Because it is often less obvious to users and the wider world, the ability to enforce quietly using qualities of the technology itself is worrisome. Technologies that lend themselves to an easy and tightly coupled expression of governmental power simply will be portable from one society to the next. It will make irrelevant the question about how firms like Google and Skype should operate outside their home countries.

This conclusion suggests that although some social gain may result from better enforcement of existing laws in free societies, the gain might be more than offset by better enforcement in societies that are less free—under repressive governments today, or anywhere in the future. If the gains and losses remain coupled, it might make sense to favor retention of generative technologies to put what law professor James Boyle has called the "Libertarian gotcha" to authoritarian regimes: if one wants technological progress and the associated eco-

nomic benefits, one must be prepared to accept some measure of social liberalization made possible with that technology.[56] Like many regimes that want to harness the benefits of the market while forgoing political liberalization, China is wrestling with this tension today.[57] In an attempt to save money and establish independence from an overseas software vendor like Microsoft, China has encouraged the adoption of GNU/Linux,[58] an operating system least amenable in its current form to appliancization because anyone can modify it and install it on a non-locked-down endpoint PC. China's attempt, therefore, represents either a misunderstanding of the key role that endpoints can play in regulation or a calculated judgment that the benefits of international technological independence outweigh the costs of less regulability.

If one objects to censorship in societies that have not developed the rule of law, one can support the maintenance of a generative core in information technology, minimizing the opportunities for some societies that wish to exploit the information revolution to discover new tools for control.

Amplification and the Lock-in of Mistakes

When a regulator makes mistakes in the way it construes or applies a law, a stronger ability to compel compliance implies a stronger ability to compel compliance with all mandates, even those that are the results of mistaken interpretations. Gaps in translation may also arise between a legal mandate and its technological manifestation. This is especially true when technological design is used as a preemptive measure. Under U.S. First Amendment doctrine, prior restraints on speech—preventing speech from occurring in the first place, rather than punishing it after the fact if indeed it is unlawful—are greatly disfavored.[59] Design features mandated to prevent speech-related behaviors, on the premise that such behaviors might turn out to be unlawful, could be thought to belong in just that category.[60] Consider the Australian Web hosting company that automatically deletes all of its clients' multimedia files every night unless it receives specific assurances up front that the files in a given directory are placed with the permission of the copyright owner or are uncopyrighted.[61]

Preemptive design may have a hard time tailoring the technical algorithms to the legal rules. Even with some ongoing human oversight, the blacklists of objectionable Web sites maintained by commercial filtering programs are consistently overbroad, erroneously placing Web sites into categories to which they do not belong.[62] For example, when the U.S. government sponsored a service to assist Iranians in overcoming Internet filtering imposed by the Iranian gov-

ernment, the U.S.-sponsored service in turn sought to filter out pornographic sites so that Iranians would not use the circumvention service to obtain pornography. The service filtered any site with "ass" in its domain name—including usembassy.state.gov, the U.S. Department of State's online portal for its own overseas missions.[63]

In the realm of copyright, whether a particular kind of copying qualifies for a fair use defense is in many instances notoriously difficult to determine ahead of time.[64] Some argue that broad attempts to embed copyright protections in technology fall short because the technology cannot easily take into account possible fair use defenses.[65] The law prohibiting the circumvention of trusted systems disregards possibilities for fair use—which might make sense, since such an exception could swallow the rule.[66] Such judgments appear to rely on the fact that the materials within a trusted system can still be found and copied in non-trusted analog formats, thus digital prohibitions are never complete.[67] The worry that a particular speech-related activity will be precluded by design is blunted when the technology merely makes the activity less convenient rather than preventing it altogether. However, if we migrate to an information ecosystem in which tethered appliances predominate, that analog safety valve will wane.

For specific injunctions, the worries about mistakes may appear weaker. A specific injunction to halt an activity or destroy its fruits issues only after an adjudication. If we move to a regime in which individuals, and not just distributors, are susceptible to impoundment remedies for digital contraband, these remedies might be applied only after the status of the contraband has been officially determined.[68] Indeed, one might think that an ability to easily recall infringing materials after the fact might make it possible to be more generous about allowing distribution in the first place—cases could proceed to final judgments rather than being functionally decided in earlier stages on the claim that continued distribution of the objectionable material would cause irreparable harm. If cats can easily be put back into bags, there can be less worry about letting them out to begin with.

However, the ability to perfectly (in the sense of thoroughly) scrub everyone's digital repositories of unlawful content may compromise the values that belie fear of prior restraints, even though the scrub would not be "prior" in fact. Preventing the copying of a work of copyrighted music stops a behavior without removing the work from the public sphere, since presumably the work is still available through authorized channels. It is a different matter to eliminate entirely a piece of digital contraband. Such elimination can make it difficult to

understand, reevaluate, or even discuss what happened and why. In ruling against a gag order at a trial, the U.S. Supreme Court worried that the order was an "immediate and irreversible sanction."[69] "If it can be said that a threat of criminal or civil sanctions after publication 'chills' speech, prior restraint 'freezes' it at least for the time."[70] Post hoc scrubs are not immediate, but they have the prospect of being permanent and irreversible—a freezing of speech that takes place after it has been uttered, and no longer just "for the time." That the speech had an initial opportunity to be broadcast may make a scrub less worrisome than if it were blocked from the start, but removing this information from the public discourse means that those who come after us will have to rely on secondary sources to make sense of its removal.

To be sure, we can think of cases where complete elimination would be ideal. These are cases in which the public interest is not implicated, and for which continued harm is thought to accrue so long as the material circulates: leaked medical records, child abuse images, and nuclear weapon designs.[71] But the number of instances in which legal judgments effecting censorship are overturned or revised—years later—counsels that an ability to thoroughly enforce bans on content makes the law too powerful and its judgments too permanent, since the material covered by the judgment would be permanently blocked from the public view. Imagine a world in which all copies of once-censored books like *Candide, The Call of the Wild,* and *Ulysses* had been permanently destroyed at the time of the censoring and could not be studied or enjoyed after subsequent decision-makers lifted the ban.[72] In a world of tethered appliances, the primary backstop against perfectly enforced mistakes would have to come from the fact that there would be different views about what to ban found among multiple sovereigns—so a particular piece of samizdat might live on in one jurisdiction even as it was made difficult to find in another.

The use of tethered appliances for surveillance may be least susceptible to an objection of mistake, since surveillance can be used to start a case rather than close it. For example, the use of cameras at traffic lights has met with some objection because of the level of thoroughness they provide—a sense of snooping simply not possible with police alone doing the watching.[73] And there are instances where the cameras report false positives.[74] However, those accused can have their day in court to explain or deny the charges inspired by the cameras' initial reviews. Moreover, since running a red light might cause an accident and result in physical harm, the cameras seem well-tailored to dealing with a true hazard, and thus less objectionable. And the mechanization of identifying violators might even make the system more fair, because the occupant of the vehi-

cle cannot earn special treatment based on individual characteristics like race, wealth, or gender. The prospects for abuse are greater when the cameras in mobile phones or the microphones of OnStar can be serendipitously repurposed for surveillance. These sensors are much more invasive and general purpose.

Bulwarks Against Government

There has been a simmering debate about the meaning of the Second Amendment to the U.S. Constitution, which concerns "the right of the people to keep and bear Arms."[75] It is not clear whether the constitutional language refers to a collective right that has to do with militias, or an individual one that could more readily be interpreted to preclude gun control legislation. At present, most reported decisions and scholarly authority favor the former interpretation, but the momentum may be shifting.[76] For our purposes, we can extract one strand from this debate without having to join it: one reason to prohibit the government's dispossession of individual firearms is to maintain the prospect that individuals could revolt against a tyrannical regime, or provide a disincentive to a regime considering going down such a path.[77] These check-on-government notions are echoed by some members of technical communities, such as those who place more faith in their own encryption to prevent secrets from being compromised than in any government guarantees of self-restraint. Such a description may unnecessarily demean the techies' worries as a form of paranoia. Translated into a more formal and precise claim, one might worry that the boundless but unnoticeable searches permitted by digital advances can be as disruptive to the equilibrium between citizen and law enforcement as any enforcement-thwarting tools such as encryption.

The equilibrium between citizens and law enforcement has crucially relied on some measure of citizen cooperation. Abuse of surveillance has traditionally been limited not simply by the conscience of those searching or by procedural rules prohibiting the introduction of illegally obtained evidence, but also by the public's own objections. If occasioned through tethered appliances, such surveillance can be undertaken almost entirely in secret, both as a general matter and for any specific search. Stuntz has explained the value of a renewed focus on physical "data mining" via group sweeps—for example, the searching of all cars near the site of a terrorist threat—and pointed out that such searches are naturally (and healthily) limited because large swaths of the public are noticeably burdened by them.[78] The public, in turn, can effectively check such government action by objecting through judicial or political processes, should the sweeps become too onerous. No such check is present in the controlled digital

environment; extensive searching can be done with no noticeable burden—indeed, without notice of any kind—on the parties searched. For example, the previously mentioned FBI use of an OnStar-like system to listen in on the occupants of a car is public knowledge only because the manufacturer chose to formally object.[79]

The rise of tethered appliances significantly reduces the number and variety of people and institutions required to apply the state's power on a mass scale. It removes a practical check on the use of that power. It diminishes a rule's ability to attain legitimacy as people choose to participate in its enforcement, or at least not stand in its way.

A government able to pressure the provider of BlackBerries could insist on surveillance of e-mails sent to and from each device.[80] And such surveillance would require few people doing the enforcement work. Traditionally, ongoing mass surveillance or control would require a large investment of resources and, in particular, people. Eavesdropping has required police willing to plant and monitor bugs; seizure of contraband has required agents willing to perform raids. Further, a great deal of routine law enforcement activity has required the cooperation of private parties, such as landlords, banks, and employers. The potential for abuse of governmental power is limited not only by whatever procedural protections are afforded in a jurisdiction that recognizes the rule of law, but also more implicitly by the decisions made by parties asked to assist. Sometimes the police refuse to fire on a crowd even if a dictator orders it, and, less dramatically, whistleblowers among a group of participating enforcers can slow down, disrupt, leak, or report on anything they perceive as abusive in a law enforcement action.[81]

Compare a citywide smoking ban that enters into effect as each proprietor acts to enforce it—under penalty for failing to do so, to be sure—with an alternative ordinance implemented by installing highly sensitive smoke detectors in every public place, wired directly to a central enforcement office. Some in favor of the ordinance may still wish to see it implemented by people rather than mechanical fiat. The latter encourages the proliferation of simple punishment-avoiding behavior that is anathema to open, participatory societies. As law professor Lior Strahilevitz points out, most laws are not self-enforcing, and a measure of the law's value and importance may be found in just how much those affected by it (including as victims) urge law enforcement to take a stand, or invoke what private rights of action they may have.[82] Strahilevitz points to laws against vice and gambling, but the idea can apply to the problems arising from technology as well. Law ought to be understood not simply by its meaning as a

text, but by the ways in which it is or is not internalized by the people it affects—whether as targets of the law, victims to be helped by it, or those charged with enforcing it.[83]

The Benefits of Tolerated Uses

A particular activity might be illegal, but in some cases those with standing to complain about it sometimes hold back on trying to stop it while they determine whether they really object. If they decide they do object, they can sue. Tim Wu calls this phenomenon "tolerated uses,"[84] and copyright infringement shows how it can work.

When Congress passed the Digital Millennium Copyright Act of 1998 (DMCA),[85] it sought to enlist certain online service providers to help stop the unauthorized spread of copyrighted material. ISPs that just routed packets for others were declared not responsible for copyright infringement taking place over their communication channels.[86] Intermediaries that hosted content—such as the CompuServe and Prodigy forums, or Internet hosting sites such as Geocities.com—had more responsibility. They would be unambiguously clear of liability for copyright infringement only if they acted expeditiously to take down infringing material once they were specifically notified of that infringement.[87]

Although many scholars have pointed out deficiencies and opportunities for abuse in this notice-and-takedown regime,[88] the scheme reflects a balance. Under the DMCA safe harbors, intermediaries have been able to provide flexible platforms that allow for a broad variety of amateur expression. For example, Geocities and others have been able to host personal home pages, precursors to the blogs of today, without fear of copyright liability should any of the home page owners post infringing material—at least so long as they act after specific notification of an infringement. Had these intermediaries stopped offering these services for fear of crushing liability under a different legal configuration, people would have had far fewer options to broadcast online: they could have either hosted content through their own personal PCs, with several incumbent shortcomings,[89] or forgone broadcasting altogether. Thanks to the incentives of notice-and-takedown, copyright holders gained a ready means of redress for the most egregious instances of copyright infringement, without chilling individual expression across the board in the process.

The DMCA legal regime supports the procrastination principle, allowing for experimentation of all sorts and later reining in excesses and abuses as they happen, rather than preventing them from the outset. Compelling copyright

holders to specifically demand takedown may seem like an unnecessary burden, but it may be helpful to them because it allows them to tolerate some facially infringing uses without forcing copyright holders to make a blanket choice between enforcement and no enforcement. Several media companies and publishers simply have not figured out whether YouTube's and others' excerpts of their material are friend or foe. Companies are not monolithic, and there can be dissenting views within a company on the matter. A company with such diverse internal voices cannot come right out and give an even temporary blessing to apparent copyright infringement. Such a blessing would cure the material in question of its unlawful character, because the infringement would then be authorized. Yet at the same time, a copyright holder may be loath to issue DMCA notices to try to get material removed each time it appears, because clips can serve a valuable promotional function.

The DMCA regime maintains a loose coupling between the law's violation and its remedy, asking publishers to step forward and affirmatively declare that they want specific material wiped out as it arises and giving publishers the luxury to accede to some uses without forcing intermediaries to assume that the copyright holder would have wanted the material to be taken down. People might make videos that include copyrighted background music or television show clips and upload them to centralized video sharing services like YouTube. But YouTube does not have to seek these clips out and take them down unless it receives a specific complaint from the copyright holder.

While requiring unprompted attempts at copyright enforcement by a firm like YouTube may not end up being unduly burdensome to the intermediary—it all depends on how its business model and technology are structured—requiring unprompted enforcement may end up precluding uses of copyrighted material to which the author or publisher actually does not object, or on which it has not yet come to a final view.[90]

Thus there may be some cases when preemptive regimes can be undesirable to the entities they are designed to help. A preemptive intervention to preclude some particular behavior actually disempowers the people who might complain about it to decide that they are willing, after all, to tolerate it. Few would choose to tolerate a murder, making it a good candidate for preemption through design, were that possible,[91] but the intricacies of the markets and business models involved in the distribution of intellectual works means that reasonable copyright holders could disagree on whether it would be a good thing to prevent certain unauthorized distributions of their works.

The generative history of the Internet shows that allowing openness to third-

party innovation from multiple corners and through multiple business models (or no business model at all) ends up producing widely adopted, socially useful applications not readily anticipated or initiated through the standard corporate production cycle.[92]

For example, in retrospect, permitting the manufacture of VCRs was a great boon to the publishers who were initially opposed to it. The entire video rental industry was not anticipated by publishers, yet it became a substantial source of revenue for them.[93] Had the Hush-A-Phones, Carterfones, and modems of Chapter Two required preapproval, or been erasable at the touch of a button the way that an EchoStar DVR of today can be killed, the decisions to permit them might have gone the other way, and AT&T would not have benefited as people found new and varied uses for their phone lines.

Some in the music, television, and movie industries are embracing cheap networks and the free flow of bits, experimenting with advertising models similar to those pioneered for free television, in which the more people who watch, the more money the publishers can make. For instance, the BBC has made a deal with the technology firm Azureus, makers of a peer-to-peer BitTorrent client that has been viewed as contraband on many university campuses and corporate networks.[94] Users of Azureus's software will now be able to download BBC television programs for free, and with authorization, reflecting both a shift in business model for the BBC and a conversion of Azureus from devil's tool to helpful distribution vehicle. BitTorrent software ensures that people upload to others as they download, which means that the BBC will be able to release its programs online without incurring the costs of a big bandwidth bill because many viewers will be downloading from fellow viewers rather than from the BBC. EMI is releasing music on iTunes without digital rights management—initially charging more for such unfettered versions.[95]

The tools that we now take as so central to the modern Internet, including the Web browser, also began and often remain on uncertain legal ground. As one surfs the Internet, it is easy to peek behind the curtain of most Web sites by asking the browser to "view source," thereby uncovering the code that generates the viewed pages. Users can click on nearly any text or graphic they see and promptly copy it to their own Web sites or save it permanently on their own PCs. The legal theories that make these activities possible are tenuous. Is it an implied license from the Web site owner? Perhaps, but what if the Web site owner has introductory text that demands that no copies like that be made?[96] Is it fair use? Perhaps. In the United States, fair use is determined by a fuzzy four-factor test that in practice rests in part on habit and custom, on people's

expectations.[97] When a technology is deployed early, those expectations are unsettled, or perhaps settled in the wrong direction, especially among judges who might be called upon to apply the law without themselves having fully experienced the technologies in question. A gap between deployment and regulatory reaction gives the economic and legal systems time to adapt, helping to ensure that doctrines like fair use are applied appropriately.

The Undesirable Collapse of Conduct and Decision Rules

Law professor Meir Dan-Cohen describes law as separately telling people how to behave and telling judges what penalties to impose should people break the law. In more general terms, he has observed that law comprises both conduct rules and decision rules.[98] There is some disconnect between the two: people may know what the law requires without fully understanding the ramifications for breaking it.[99] This division—what he calls an "acoustic separation"—can be helpful: a law can threaten a tough penalty in order to ensure that people obey it, but then later show unadvertised mercy to those who break it.[100] If the mercy is not telegraphed ahead of time, people will be more likely to follow the law, while still benefiting from a lesser penalty if they break it and have an excuse to offer, such as duress.

Perfect enforcement collapses the public understanding of the law with its application, eliminating a useful interface between the law's terms and its application. Part of what makes us human are the choices that we make every day about what counts as right and wrong, and whether to give in to temptations that we believe to be wrong. In a completely monitored and controlled environment, those choices vanish. One cannot tell whether one's behavior is an expression of character or is merely compelled by immediate circumstance.

Of course, it may be difficult to embrace one's right to flout the law if the flouting entails a gross violation of the rights of another. Few would uphold the freedom of someone to murder as "part of what makes us human." So we might try to categorize the most common lawbreaking behaviors online and see how often they relate to "merely" speech-related wrongs rather than worse transgressions. This is just the sort of calculus by which prior restraints are disfavored especially when they attach to speech, rather than when they are used to prevent lawbreaking behaviors such as those that lead to physical harm. If most of the abuses sought to be prevented are well addressed through post hoc remedies, and if they might be adequately discovered through existing law enforcement mechanisms, one should disfavor perfect enforcement to preempt them. At the

very least, the prospect of abuse of powerful, asymmetric law enforcement tools reminds us that there is a balance to be struck rather than an unmitigated good in perfect enforcement.

WEB 2.0 AND THE END OF GENERATIVITY

The situation for online copyright illustrates that for perfect enforcement to work, generative alternatives must not be widely available.[101] In 2007, the movie industry and technology makers unveiled a copy protection scheme for new high-definition DVDs to correct the flaws in the technical protection measures applied to regular DVDs over a decade earlier. The new system was compromised just as quickly; instructions quickly circulated describing how PC users could disable the copy protection on HD-DVDs.[102] So long as the generative PC remains at the center of the modern information ecosystem, the ability to deploy trusted systems with restrictions that interfere with user expectations is severely limited: tighten a screw too much, and it will become stripped.

So could the generative PC ever really disappear? As David Post wrote in response to a law review article that was a precursor to this book, "a grid of 400 million open PCs is not less generative than a grid of 400 million open PCs and 500 million locked-down TiVos."[103] Users might shift some of their activities to tethered appliances in response to the security threats described in Chapter Three, and they might even find themselves using locked-down PCs at work or in libraries and Internet cafés. But why would they abandon the generative PC at home? The prospect may be found in "Web 2.0." As mentioned earlier, in part this label refers to generativity at the content layer, on sites like Wikipedia and Flickr, where content is driven by users.[104] But it also refers to something far more technical—a way of building Web sites so that users feel less like they are looking at Web pages and more like they are using applications on their very own PCs.[105] New online map services let users click to grasp a map section and move it around; new Internet mail services let users treat their online e-mail repositories as if they were located on their PCs. Many of these technologies might be thought of as technologically generative because they provide hooks for developers from one Web site to draw upon the content and functionality of another—at least if the one lending the material consents.[106]

Yet the features that make tethered appliances worrisome—that they are less generative and that they can be so quickly and effectively regulated—apply

with equal force to the software that migrates to become a service offered over the Internet. Consider Google's popular map service. It is not only highly useful to end users; it also has an open API (application programming interface) to its map data,[107] which means that a third-party Web site creator can start with a mere list of street addresses and immediately produce on her site a Google Map with a digital push-pin at each address.[108] This allows any number of "mash-ups" to be made, combining Google Maps with third-party geographic datasets. Internet developers are using the Google Maps API to create Web sites that find and map the nearest Starbucks, create and measure running routes, pinpoint the locations of traffic light cameras, and collate candidates on dating sites to produce instant displays of where one's best matches can be found.[109]

Because it allows coders access to its map data and functionality, Google's mapping service is generative. But it is also contingent: Google assigns each Web developer a key and reserves the right to revoke that key at any time, for any reason—or to terminate the whole Google Maps service.[110] It is certainly understandable that Google, in choosing to make a generative service out of something in which it has invested heavily, would want to control it. But this puts within the control of Google, and anyone who can regulate Google, all downstream uses of Google Maps—and maps in general, to the extent that Google Maps' popularity means other mapping services will fail or never be built.

Software built on open APIs that can be withdrawn is much more precarious than software built under the old PC model, where users with Windows could be expected to have Windows for months or years at a time, whether or not Microsoft wanted them to keep it. To the extent that we find ourselves primarily using a particular online service, whether to store our documents, photos, or buddy lists, we may find switching to a new service more difficult, as the data is no longer on our PCs in a format that other software can read. This disconnect can make it more difficult for third parties to write software that interacts with other software, such as desktop search engines that can currently paw through everything on a PC in order to give us a unified search across a hard drive. Sites may also limit functionality that the user expects or assumes will be available. In 2007, for example, MySpace asked one of its most popular users to remove from her page a piece of music promotion software that was developed by an outside company. She was using it instead of MySpace's own code.[111] Google unexpectedly closed its unsuccessful Google Video purchasing service and remotely disabled users' access to content they had purchased; after an outcry, Google offered limited refunds instead of restoring access to the videos.[112]

Continuous Internet access thus is not only facilitating the rise of appliances and PCs that can phone home and be reconfigured by their vendors at any moment. It is also allowing a wholesale shift in code and activities from endpoint PCs to the Web. There are many functional advantages to this, at least so long as one's Internet connection does not fail. When users can read and compose e-mail online, their inboxes and outboxes await no matter whose machines they borrow—or what operating system the machines have—so long as they have a standard browser. It is just a matter of getting to the right Web site and logging in. We are beginning to be able to use the Web to do word processing, spreadsheet analyses, indeed, nearly anything we might want to do.

Once the endpoint is consigned to hosting only a browser, with new features limited to those added on the other end of the browser's window, consumer demand for generative PCs can yield to demand for boxes that look like PCs but instead offer only that browser. Then, as with tethered appliances, when Web 2.0 services change their offerings, the user may have no ability to keep using an older version, as one might do with software that stops being actively made available.

This is an unfortunate transformation. It is a mistake to think of the Web browser as the apex of the PC's evolution, especially as new peer-to-peer applications show that PCs can be used to ease network traffic congestion and to allow people directly to interact in new ways.[113] Just as those applications are beginning to show promise—whether as ad hoc networks that PCs can create among each other in the absence of connectivity to an ISP, or as distributed processing and storage devices that could apply wasted computing cycles to faraway computational problems[114]—there is less reason for those shopping for a PC to factor generative capacity into a short-term purchasing decision. As a 2007 *Wall Street Journal* headline put it: "'Dumb terminals can be a smart move': Computing devices lack extras but offer security, cost savings."[115]

* * *

Generative networks like the Internet can be partially controlled, and there is important work to be done to enumerate the ways in which governments try to censor the Net.[116] But the key move to watch is a sea change in control over the endpoint: lock down the device, and network censorship and control can be extraordinarily reinforced. The prospect of tethered appliances and software as service permits major regulatory intrusions to be implemented as minor technical adjustments to code or requests to service providers. Generative technologies ought to be given wide latitude to find a variety of uses—including

ones that encroach upon other interests. These encroachments may be undesirable, but they may also create opportunities to reconceptualize the rights underlying the threatened traditional markets and business models. An information technology environment capable of recursive innovation[117] in the realms of business, art, and culture will best thrive with continued regulatory forbearance, recognizing that the disruption occasioned by generative information technology often amounts to a long-term gain even as it causes a short-term threat to some powerful and legitimate interests.

The generative spirit allows for all sorts of software to be built, and all sorts of content to be exchanged, without anticipating what markets want—or what level of harm can arise. The development of much software today, and thus of the generative services facilitated at the content layer of the Internet, is undertaken by disparate groups, often not acting in concert, whose work can become greater than the sum of its parts because it is not funneled through a single vendor's development cycle.[118]

The keys to maintaining a generative system are to ensure its internal security without resorting to lockdown, and to find ways to enable enough enforcement against its undesirable uses without requiring a system of perfect enforcement. The next chapters explore how some enterprises that are generative at the content level have managed to remain productive without requiring extensive lockdown or external regulation, and apply those lessons to the future of the Internet.