



Discovery, Rediscovery, and Open Access, Parts 1 and 2

Citation

Suber, Peter, "Discovery, Rediscovery, and Open Access, Parts 1 and 2" SPARC Open Access Newsletter, August and September 2010

Link

<https://dash.harvard.edu/handle/1/42718524>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Attribution (CC-BY), as set forth at <https://creativecommons.org/licenses/by/4.0/>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

Discovery, Rediscovery, and Open Access

Author's note. In August and September 2010, I published a two-part essay titled, “Discovery, rediscovery, and open access.” Here I put the two parts together for the first time. I’ve fixed dead links and changed the wording in a few places to improve clarity.

[Discovery, rediscovery, and open access. Part 1.](#)

SPARC Open Access Newsletter, issue #148

August 2, 2010

by Peter Suber

In 1979, William Garvey made a remarkable claim: “... in some disciplines, it is easier to repeat an experiment than it is to determine that the experiment has already been done.” (See W.D. Garvey, *Communication: The essence of science*, Pergamon Press, Oxford 1979, p. 8.)

Garvey was talking about research in the era of print, and we’d like to think that digital technologies have changed the picture. But Garvey’s thesis is not false today. It’s just true less often than it was in 1979.

Of course digitizing research makes it easier to find. But when finding it is still hard (because search tools are weak or access barriers block crawlers) or when retrieval is hard (because the work is paywalled) or when the original experiment is particularly easy to repeat, then repeating the experiment can still be the path of least resistance.

The same is true when we move beyond mere digitization to online distribution. Search will be easier, but may still fail. For many people, retrieval will still be costly or impossible.

The same is true when we move beyond mere online distribution to online and open access (OA) distribution. When the original research is OA, then if we can find it, we can retrieve it. But a hard search could still be harder than an easy experiment — though not very often.

Of course even when the original research is findable and retrievable, the researchers might have been vague or coy when writing it up. (“Decant the phlogiston...” or “The results, according to our [homegrown, closed-source] software....”). But the Garvey problem is about access failures, not author failures. Hence, we needn’t consider the case — which can arise in an OA world — in which we can find and retrieve the original research but must still repeat the work in order to make up for deficiencies in what we retrieved.

Our best hope for solving the Garvey problem is the combination of ubiquitous OA and powerful search, and fortunately we’re making good progress on both fronts. The scope of the Garvey problem is shrinking. The optimistic take is that it won’t be long before the Garvey problem is limited to known, simple truths that are easier to discover in the world than to rediscover by search. (For the pessimistic take, see Part 2 next month.)

This class of simple truths may be small and shrinking, but it’s non-empty. For example, I have a thermometer outside my kitchen window, but I wear bifocals and can’t read it from across the room. On some days, when I’m drinking my morning coffee at my laptop, it’s easier to Google the local temperature than walk across the room. On other days, it’s easier to walk across the room. A key variable is whether Bonny, my excitable 100-pound Labrador Retriever, is hovering between me and the window waiting for me to stand up.

Of course when the Garvey problem arises only for this class of truths, it’s no longer a serious obstacle. As time goes on, then, we might be tempted to blur the distinction between the cases where there is no Garvey problem and the cases where the Garvey problem is no obstacle — that is, between the cases when searching is easier than repeating the original experiment and the cases when searching is harder but still easy. But we should remember that we reach this plateau of low-barrier access to knowledge not just in the rare cases when discovery and rediscovery are both easy. We also reach it in the more plentiful cases when just one of them is easy. Hence, it remains important to distinguish the speed bump of search from the speed bump of repeating the original experiment or observations. The two bumps are equally low only for a special

subset of simple truths. But we only need one to be low in order to accelerate research.

Where we can't make empirical discovery easier, at least we can make look-up or rediscovery easier. Step One in eliminating the Garvey problem is to discover and record a piece of knowledge, or a claimed piece of knowledge. If we get this far, then in principle others can find it without having to repeat the original work. Step Two is to make finding easier than repeating. Hence, it takes a village to shrink the scope of the Garvey problem: the original discoverers and recorders, the larger community of refiners and confirmers, and the army of cooperators developing the recent critical improvements to our access system: digitization, online distribution, strengthening search, and spreading OA.

(For more on why OA isn't enough, and why search must complement OA, see [SOAN for July 2005](#). For more on why search isn't enough, and why OA must complement search, see [SOAN for December 2005](#).)

Solving the Garvey problem doesn't mean that we'll always make use of what we know. But it gives us a fighting chance. It means that we're moving beyond mere preservation to a useful degree of findability. Findability may exist in a spectrum of degrees from 0 to 1. But we pass an important threshold when the ease of findability through an effective access system surpasses the ease of discoverability through fresh thinking and experimentation. Insofar as knowledge is worth accessing, knowing, or using, it's urgent to pass this threshold and urgent to remain above it. It's just as urgent to keep working to lower any access barriers that might remain.

Where we haven't solved the Garvey problem, we're allowing access barriers to render some known and recorded truths essentially useless. Where we've solved the Garvey problem but haven't kept working to lower the barriers to findability and retrievability as much as we can, we're allowing access barriers to render some known and recorded truths needlessly essentially invisible.

(1) For a given research question, I might face a Garvey problem when you would not. For example, you might have good access to the internet or a print library

when I don't. Or I might have good access to lab equipment when you don't. (Further, one of us might face a lab problem and the other a "Lab" problem.)

Similarly, some research questions raise Garvey problems when others do not. Some searches are so difficult — reading a handwritten manuscript locked in the Vatican — that we'd prefer to redo the original work, if only we could. This is the kind of problem that digitization, OA, and search can solve, even if they haven't yet solved them all.

Conversely, some experiments are too expensive or dangerous to repeat. Does lunar soil contain more silicon than iron? What happens when two 3.5 trillion electron-volt proton beams collide? Is thalidomide safe for pregnant women? If only we could answer our question through a difficult search, perhaps by traveling across the globe to read a unique handwritten manuscript, we'd prefer to do that than to redo the experiment.

OA helps in both cases, either by making difficult searches easier or by making the repetition of difficult experiments unnecessary.

For more on the second family of cases, see [Richard Poynder's case study of CERN's Large Hadron Collider](#) (August 2008), showing that [big science intrinsically carries big incentives for OA](#):

"Ten or 20 years ago we might have been able to repeat an experiment," says [CERN's Rolf-Dieter Heuer]. "They were simpler, cheaper and on a smaller scale. Today that is not the case. So if we need to re-evaluate the data we collect to test a new theory, or adjust it to a new development, we are going to have to be able to reuse it. That means we are going to need to save it as open data." ...Openness is not an issue for data alone, however. The research papers produced from the LHC experiments will also have to be open....Because science is a cumulative process, the greater the number of people who can access research, critique it, check it against the underlying data and then build on it, the sooner new solutions and theories will emerge. And as "Big Science" projects like the LHC become the norm,

the need for openness will be even greater because the larger the project, the more complex the task, and the greater the need for collaboration....Certainly, if the public is asked to fund further multi-billion-pound projects like the [Large Hadron Collider], there will be growing pressure on scientists to maximise the value of the data they generate — and that will require greater openness.

Banks too big to fail can threaten the economy. But experiments too big to repeat needn't threaten research if we open up the process, data, software, code, and results in order to minimize the need for repetition. We will still need repetition to test reproducibility, but not to solve the Garvey problem.

Experiments too big to repeat needn't face Garvey problems. No matter how difficult it is to find the original results, provided they still exist, finding them will be easier than redoing the original work. But insofar as we tolerate access barriers to those results, we undermine our own enormous investment in the original research. It's penny wise and pound foolish to fund an expensive experiment and make the results expensive for subsequent researchers to find and retrieve.

If this is true of single experiments that are individually expensive, it's also true of portfolios of experiments that are collectively expensive. One of the compelling rationales for a funder OA mandate is that the large investment in the funder's research budget should not be undermined by needless access barriers to the results. Many funders turn to OA to solve "big portfolio" and "big science" problems at the same time. The NIH research budget is more than the GDP of 140 nations. When taxpayers devote that kind of money to research, they can maximize the return on their investment by ensuring that the results are available to all who can build on them. In addition, the cost of an NIH-funded research project can be hundreds or even thousands of times greater than the cost of publication. To allow its results to be held hostage by publishers is the same mistake on a different scale as spending billions on a Large Hadron Collider and locking up the results in paywalled publications.

(2) Not all the literature we want to find, retrieve, and read should be called "knowledge". We want access to serious proposals for knowledge even if we're still

evaluating them and debating their merits. We want access even to knowledge claims that turn out to be false or one-sided. We want access to datasets even when they are small, uninterpreted, or methodologically flawed. We want access to observations that were true when they were made but time-bound and variable (like Garvey's own observation in 1979). We want access to all the data, evidence, arguments, and algorithms that help us decide what to call “knowledge”, not just to the results that we agree to call “knowledge”. If access depended on the *outcome* of debate and inquiry, then access could not *contribute* to debate and inquiry.

We don't have a good name for this category larger than knowledge, but here I'll just call it “research”. Research includes knowledge, knowledge claims and proposals, hypotheses and conjectures, data, algorithms, debate, evaluation, criticism, dissent, interpretation, summary, and review.

Recapitulating an experiment or observation can be trivial compared to recapitulating a generations-long and culture-wide (or cultures-wide) debate. Large-scale debates, like large-scale experiments, are too big to repeat. Like large-scale experiments, they are generally immune to Garvey problems. Even if finding and retrieving all the relevant bits is extraordinarily difficult, it's usually easier than recapitulating the original process. It doesn't follow that access barriers are low, merely that they are lower than repetition barriers. Access can be formidably expensive and difficult, not only because of price and permission barriers, but because of the nearly opposite problems of information overload and inadequate filtering and search.

In these ways, research in the wider sense, and nearly all research in the humanities, is like big science. It's too big to repeat or recapitulate and carries intrinsic incentives for OA. Leaving access barriers any higher than necessary means slowing the process of inquiry and wasting more time, effort, and resources than necessary.

Postscript. This essay stands on its own, but it's also Part 1 of a longer piece. I'm hoping to publish Part 2 next month. In Part 2 I'll look at (1) when we *want* to repeat the original work, to test its reproducibility, and (2) when the rediscovery

of older knowledge is made difficult not so much by access barriers as by ignorance, indifference, taboos, and Dark Ages 2.0.

[Discovery, rediscovery, and open access. Part 2.](#)

SPARC Open Access Newsletter, issue #149

September 2, 2010

by Peter Suber

In Part 1 of this essay (published in [SOAN for August 2010](#)) I sketched some ways in which the growth of OA modified William Garvey's 1979 observation that "in some disciplines, it is easier to repeat an experiment than it is to determine that the experiment has already been done."

Here I'd like to connect OA with three variations on Garvey's theme. Garvey focused on cases in which redoing past work is undesirable but easier than looking up the original results. The problem to solve or work around is a dysfunctional access system. Sometimes, however, we positively want to redo past work. The problem to solve is that the original results are untested or unconfirmed, not inaccessible. Sometimes we redo past work inadvertently. The problem to solve is our near-sighted review of past literature. Sometimes redoing past work and looking it up are both undesirable. The problem to solve is that we've allowed knowledge to become taboo and replaced curiosity with a defensive preference for what we already believe. Anything is easier than looking up past work or redoing it. All literature reviews are near-sighted. The problem to solve lies in us, our fears and complacency, or in our predecessors, who might have broken the access system, burned the books, or created a culture in which inquiry is stigmatized as disloyal and harmful to party, profits, or faith.

(1) In classic Garvey cases, redoing past work is a distant second-best to reading the original results. But when the original results have yet to be replicated, then reading them, without more, is a distant second-best to redoing the work itself. In Garvey cases, lack of access forces our hand and we redo experiments even if they have already been replicated and confirmed. In cases of untested results, lack of

replication forces our hand and we redo experiments even if access to the original records is easy and open.

This Garvey variation is compatible with good access to past results. We'll want to replicate untested results even when we don't have good access to the original records. But good access doesn't make the need go away. It even facilitates the work, especially when (as Victoria Stodden and others argue) OA to texts is complemented by open data and open software, all under open licenses.

See the "[oa.reproducibility](#)" tag library for the Open Access Tracking Project.

In short, when we want to redo research, OA helps us do so. When we'd rather look up past research than redo it, OA helps us there too. OA doesn't remove or reduce the need to test untested results, but it does remove the need to work around a dysfunctional access system.

We generally talk about reproducibility only in the experimental sciences, where conclusions stand or fall on the ability of researchers elsewhere to repeat the same experiment and get the same results. But if we abstract a bit, we can see that scholars in the humanities find analogous value in revisiting older inquiries. They're not trying to reproduce them, or test their results empirically, but to reopen older — sometimes ancient — questions, criticize or reinterpret previous answers, and continue the conversation. OA facilitates both the empirical and reflective sorts of repetition, just as it helps us avoid repetition when we simply want to read the original literature and move on.

(2) Another Garvey variation arises when researchers should read earlier results but fail to do so. They repeat past work, not to test it or to work around a broken access system, but because they think it hasn't already been done.

In 1964, John Martyn estimated that about 9% of UK research projects could have saved about 10% of their budgets if they had known about relevant prior work before starting their own. Making good use of a good access system could have saved the UK national research budget about 0.9% (10% of 9%) per year. Using 1964 figures for the total UK research budget (640 million pounds) and average UK

researcher salary (8k pounds), Martyn estimated that making research accessible and actually accessing it would have saved the country enough money to hire 750 full-time researchers. Surely, he speculates, some of those researchers could have been put to work doing literature searches to prevent the costly and unwanted sorts of duplication.

See John Martyn, [Unintentional Duplication of Research](#), *New Scientist*, February 6, 1964.

I haven't seen an update to Martyn's 1964 calculations. (If you have, I'd welcome a pointer.) In 2001, Isaac Ginsburg reprised the issue without redoing the math. While Martyn focused on cases when the failure to find relevant past work was due to carelessness, Ginsburg recognized at least two additional causes: a dogmatic belief that past work couldn't be relevant or accurate, and an honest miss because the older and newer research used different terminology.

See Isaac Ginsburg, [The Disregard Syndrome: A Menace to Honest Science?](#) *The Scientist*, December 10, 2001. (Thanks to Eugene Garfield for bringing Martyn and Ginsburg's work to my attention.)

We can add one more possibility: the problem of inadvertent duplication, like the classic Garvey problem, can be caused by a broken access system.

Like Garvey's original observation, Martyn's calculation was made in the age of print and would have to be modified for the age of the internet and OA. On the one hand, OA should reduce the number of projects that fail to discover relevant prior work in time to prevent inadvertent duplication. On the other hand, the total savings from enhanced access wouldn't have to be as large to pay for the literature searches needed to maximize those savings. How this nets out I leave as an exercise for the reader.

As in the previous variation, inadvertently repeating past work in the belief that it hasn't been done is compatible with good access to the literature. Good access should reduce inadvertent duplication but doesn't always do so. Accessible literature isn't always accessed literature (a problem that will reappear below).

Hence, OA can only solve part of the Martyn-Ginsburg problem. If the problem continues in the age of OA, it will tend to highlight problems with researcher care, skill, or will.

In his 1974 essay on cargo cult science, Richard Feynman argued that even when researchers know about relevant prior work and want to replicate it, they often feel pressure to bypass replication in favor of new results, especially when they're using expensive equipment and need new results to keep the funds flowing.

See Richard Feynman, [Cargo Cult Science](#), 1974.

Martyn argued in effect that negligence works as well as access barriers to trigger the Garvey problem. Feynman argued in effect that the reward system can deter replication even when researchers understand its value and want to realize it. For Martyn, neglect can trigger the undesirable sort of repetition, while for Feynman interest can trigger neglect of the desirable sort.

(3) Bear with me while I describe a third Garvey variation in which the barriers to looking up past work, and to redoing it, are more cultural than technical. It's an extension of the Martyn/Ginsburg variation in which researcher negligence becomes cultural neglect. It's an extension of the Feynman variation in which economic interests buttress the careless kind of neglect with a more calculating kind of neglect. The problems created by this third variation are always mitigated by OA, but sometimes aggravated by it, and (therefore) sometimes both. Since it hasn't yet materialized in its extreme form, let me cast myself as a crank in order to paint the picture.

In America, a large and influential subset of the population fights fiercely for the right to own guns. This is hard to explain to the rest of the world. Part of the explanation is that many Americans fear the day when citizens will have to fight for their freedoms against their own elected, oppressive government. I don't share that fear. But I do sometimes feel another fear which is equally hard to explain to the rest of the world. I fear that a wave of barbarism will overtake civilization. If the barbarians don't burn the recorded knowledge and reflection of centuries, they will at least neglect it and cause others to neglect it.

Preservation is only a small part of the solution, since preserving knowledge only saves it from loss, not from neglect. Education is a larger part of the solution but still only a small part, since we only have a 10-20 year window in which to try our miserable best to do what formal education can do. Even if formal education were magically effective, the rapid growth of research means that with every passing year the fraction we can fit into that 10-20 year window becomes a smaller and smaller percentage of the whole. We've long since passed the point when we could integrate the bulk of human knowledge, thought, and culture into an individual life. But integration is the only solution that is robust in the face of neglect. Insofar as we can integrate our cultural inheritance into our lives, we can make use of what is useful, benefit from what is beneficial, and avoid the need to rediscover what our predecessors had already discovered.

My fear applies to settled knowledge, however you conceive that, as well as to inquiries and debates that haven't yet settled into knowledge, and to clarity in framing our questions, hypotheses, methods, and arguments. All these achievements are hard-won, and were generally won against the indifference or hostility of huge numbers of people. The background indifference and hostility do not make knowledge, inquiry, and clarity intrinsically more valuable. But they should make us pause to appreciate the difficulty of recovering them if we allow them to slip from our grasp.

We don't have to highlight indifference and hostility, which are obstacles that vary from time to time and place to place. Instead we can simply note that knowledge advances through ingenious insights, lucky accidents, trial and error, and painstaking observation. All of these are difficult or rare. Regardless of where we locate the obstacles, if knowledge is worth having, then it's worth keeping and keeping accessible. Betting that we could easily recover it would be like betting against entropy. The local exceptions to entropy are notable and thrilling, but they are still exceptions.

Last month [Newsweek declared](#) US education to be the 11th best in the world. That's not a proud result for the world's wealthiest nation but it's a respectable showing and even an improvement over some earlier surveys. Nevertheless, one

month before *Newsweek* released its results, a [Marist poll](#) found that 26% of Americans didn't know that their country won its independence from Great Britain. One week after the *Newsweek* results, a [Pew poll](#) found that 18% of Americans believed that Barack Obama is Muslim.

Three months before the *Newsweek* results, [Texas approved](#) a new history curriculum reducing the role of Thomas Jefferson and the doctrine of the separation of church and state, enlarging the role of Joseph McCarthy and Phyllis Schlafly, and [renaming the slave trade](#) the "Atlantic triangular trade".

Conservapedia, the self-described "Trustworthy Encyclopedia", is extending the fundamentalist war against Darwin to a war against Einstein. Conservapedia contributors [object that the theory of relativity](#) is "heavily promoted by liberals who like its encouragement of relativism" and have painstakingly collected 30 "[counterexamples to relativity](#)", including "the action-at-a-distance by Jesus, described in John 4:46-54."

Those are a few recent examples from one country, my country. I'm particularly distressed by what's happening in my country, but you probably have choice local examples no matter where you live. The problem isn't limited to failures of education, which lead to innocent ignorance by people who might prefer to know. It extends to failures of honesty, which lead to cynical deception and FUD by people who profit from the ignorance and uncertainty of others.

On the rise of deliberate disinformation and FUD, see Robert Proctor and Londa Schiebinger, [Agnotology](#) (Stanford University Press, 2008); Michael Specter, [Denialism](#) (Penguin, 2009); and Naomi Oreskes and Erik Conway, [Merchants of Doubt](#) (Bloomsbury, 2010).

When we can tell the difference between innocent ignorance and cynical misrepresentation, we needn't collapse the distinction. (I've often argued that the OA movement struggles against both.) But for my purposes here, they pull in the same direction. With effort we can reverse slippage in auto safety or water quality. But slippage in educating the next generation creates a downward spiral in which

each new generation is less able and less effective than the last. (How much of your time and energy do you spend swimming against that downward spiral?)

I'm not saying that this fear motivates all OA advocates. In fact, I know it doesn't. I'm not even saying that it motivates me much. It comes and goes like a bad dream. It's one thread in a braided rope of motivations, alongside the desire to accelerate research, multiply the benefits of research, and facilitate the translation of research into new medicines, useful technologies, solved problems, informed decisions, and marvelous understandings.

The chief problem with the barbarism I sometimes fear is not the loss of peace, prosperity, or due process, serious as those are. It's the loss of knowledge and the sapping of serious research and inquiry. I fear that most of what we have learned over the past few millennia is at risk of neglect, if not outright denial, that over time it may become unintelligible or taboo, and that most of what is known and worth knowing will have to be rediscovered the hard way.

The job of rediscovery might go faster than it did the first time around if some folks had access to parts of the literature that hadn't been lost, were motivated to understand it, and were willing to stay up nights studying. Or the job might go more slowly if the culture supporting curiosity and knowledge against ignorance and superstition took even longer to get a foothold than it took the last time.

Imagine that the recorded knowledge of humanity hadn't been destroyed but only suppressed, forgotten, or stigmatized as obsolete, harmful, elitist, foreign, or contrary to faith. Imagine a New Dark Age lasting for centuries. Imagine the New Dark Age slowly coming to an end as people slowly changed their attitudes toward previously recorded knowledge and research. Imagine them gradually beginning to relearn it, teach it, put it to use. We can even imagine them trying to distinguish the parts that were still true, wise, useful, or promising from the parts that only seemed so to their past proponents, and reviving serious inquiry at the borders of what they were able to understand. We could call that a New Enlightenment.

Sometimes I hope that OA is preparing us for a kind of New Enlightenment. This is a hope, not a prediction. At least we can say that OA (plus mass digitization and search) is putting more new knowledge and more old knowledge within easy reach than we've ever had within easy reach. We're creating an opportunity to mine or assimilate huge veins of knowledge that we haven't previously mined or assimilated. Today we're waking up, fitfully, to this opportunity. I know what this is like, and work on it every day, but I can't help comparing it to something I don't know and can only imagine: waking up from a New Dark Age to face the challenge of excavating the long-buried libraries of our ancestors.

The opportunity is similar to what our predecessors faced in the early days of print: "A man born in 1453, the year of the fall of Constantinople, could look back from his fiftieth year on a lifetime in which about eight million books had been printed, more perhaps than all the scribes of Europe had produced since Constantine founded his city in A.D. 330." (Michael Clapham, "Printing," in Charles Singer et al., eds., *A History of Technology*, vol. 3, *From the Renaissance to the Industrial Revolution*, Oxford University Press, 1957. p. 37; quoted by Elizabeth Eisenstein, *The Printing Revolution in Early Modern Europe*, Cambridge University Press, second edition, 2005, p. 15.)

Sometimes, however, I fear that sudden access to sudden immensity won't stimulate a New Enlightenment so much as a cultured helplessness in the face of an infinitely rising learning curve and intractable disagreement. This is a fear, not a prediction. But it isn't a groundless fear. On the contrary, it has at least four grounds. First is our inability to master the dizzyingly large and fast-growing literature on any given topic, a problem already serious in the age of print.

Second is our temptation to cherry-pick from the same immense and fast-growing literature to support our pet conclusions. To our credit, we still see this kind of cherry-picking as a form of rationalization. But we're starting to rationalize it. We tell ourselves that cherry-picking is more defensible the closer it gets to being unavoidable, and we suspect it's becoming unavoidable. Even if we know abstractly that it's a method of fortifying bias, we also know that it succeeds in fortifying bias. It may not support inquiry, but it supports sincere beliefs of all kinds, including the false kind. It's a temptation we're failing to resist. It fixes

beliefs and polarizes debates even today when we're still ready to acknowledge that we could in principle do better.

Third is our tendency to read only the sorts of people who think like us, from familiarity, vanity, or fear of uncertainty. See the large literature on this problem from Cass Sunstein's "Daily Me" in [Republic.com](#) (Princeton University Press, 2002) to Ethan Zuckerman's [TED talk](#) (July 2010).

The fourth is what I've called the Meno problem, after the Platonic dialog (Plato, *Meno*, 80.d) in which Meno asks Socrates why he should bother inquiring for truth at all. Meno reasons that either he already knows the truth or he wouldn't recognize it if stumbled across it. Socrates called this a trick argument, but he took it seriously and concluded that knowledge is closer to recollection than discovery. You don't have to find this plausible in every domain to see that it sheds light on why we feel so helpless when set adrift on a sea of conflicting claims and asked to figure out who's right. If we didn't bring clues to answers with us, then anything could be a clue to an answer. Perhaps we don't literally need to know the truth in order to recognize the truth, but we do need to agree on methods of evaluation before we can evaluate disagreements, at least if we want to do more than just create new disagreements.

(Here are a few places where I've previously discussed the Meno problem: [one](#), [two](#), [three](#), [four](#).)

It's hard to quantify the zealous energy we are putting into digitizing print literature for easier and wider access. If you have an internet connection today, you have access to more full-text books and articles than the average academic library has on its shelves, and the disparity is growing fast. In obvious ways, this is the opposite of the world of *Fahrenheit 451*, in which the same zealous energy was put into book burning. Instead of small bands of decrepit elders huddling in the woods to preserve a dwindling handful of books, networked people by the billions have access to millions of books from their desktops, laptops, or palms.

But as Mark Twain said, the person who doesn't read has no advantage over the person who can't read. Unprecedented access can trigger unprecedented learning.

Or it can trigger reactionary neglect. It can bring unprecedented learning to those inclined to learn and voluntary provincialism to those inclined to cocoon themselves from confusion or fear. It can even bring voluntary provincialism to courageous cosmopolitans who labor under a learned specialization, laser-like focus, and lack of free time.

As access increases, will we rediscover a lode of knowledge and wisdom previously overlooked? Or will we overlook a lode of knowledge and wisdom previously discovered? I don't know. But I worry about this more than I worry about black helicopters.

I worry not because the darkest outcome is the most likely. I worry because the most likely outcome, today, seems to be an extension of what we're experiencing today: the coexistence of a New Enlightenment and New Dark Age in some unpredictable and fluctuating proportion. If so, then tearing down access barriers to knowledge will not eliminate ideological disagreement or invincible ignorance. On the contrary, it could amplify the confusing profusion of knowledge claims and trigger the reflexive cling to comfortable certitudes.

Rapid progress on digitization, search, and OA reduces the Garvey problem and reduces the risk of the primary form of the New Dark Age problem: the problem of inaccessible knowledge. That's all to the good. But they leave us vulnerable to secondary forms of the New Dark Age problem: the problems of neglected or unaccessed knowledge, voluntary provincialism, self-righteous self-stultification, arguably unavoidable cherry-picking, and bottomless disagreement. That's a recurring bad dream. The best ground for hope is a recommendation that too often goes without saying. OA is necessary but not sufficient for an enlightening spread of already-discovered, already-recorded knowledge. What we must find elsewhere, to supplement mere access, is the commitment to care more for knowledge than tribal mythology or interest, and the commitment to pass on that commitment to subsequent generations.

I've admitted that I'm pessimistic enough to worry. But I'm optimistic enough to work for the partial solution we call OA. OA is a precondition of some enlightening futures and a potential catalyst of others. At worst, OA (plus mass

digitization and search) is a contributing cause to some of the secondary versions of the New Dark Age problem. But in each of those dark scenarios, OA remains a mitigating factor and ground for hope. Access to abundance may frighten some people to retreat from the risks of inquiry to the safety of certitude. But access itself preserves the permanent possibility of reviving inquiry and climbing back out of any cognitive hole we might dig for ourselves. That's a reason to work for it.

That sort of optimism is already qualified. But I'll qualify it further. I still affirm the position I took in a 2006 [blog post](#):

I'm not so optimistic as to think that simply making primary science easily available online will do much to foster scientific literacy and scientific knowledge among non-scientists, let alone convert creationists to evolutionists. Easy access completes the puzzle when there is antecedent interest and background, and we need help from teachers, journalists, and politicians to create that interest and background. For the same reason, however, I'm not so pessimistic as to think that OA will make no difference. There are two mistakes to avoid here. One is to think that OA has no role to play in helping non-scientists understand science. We can call this the [Royal Society mistake](#), after the RS's recent report [May 2006]...on educating lay readers about science that doesn't even mention OA. The other mistake is to think that the overriding purpose of OA is to educate lay readers. No OA advocates believe this, but some publisher-opponents of OA either believe it or pretend to believe it in order to set it up as a straw man and knock it down....To avoid both mistakes we have to accept that the problem and solution are both complicated. OA will play a role in public education about science — it's neither irrelevant nor sufficient — and the size of that role is up to all of us.

I believe that curiosity and knowledge-seeking are fundamental and ineradicable. But the reminder does little to boost my optimism. The same evidence that makes me worry about a New Dark Age makes me worry that curiosity and knowledge-seeking are offset by equal and opposite forces. Suspicion of curiosity, aversion to knowledge-seeking, and the countervailing demands of interest,

tribalism, and defensive dogmatism are at least as strong. When we're seeking knowledge, OA helps us in every way. But when we're indulging our countervailing interests, OA is either an annoying complexity or an aid to cherry-picking. OA is revolutionizing inquiry, but we can't expect it to change our fundamental motives for undertaking, resisting, or distorting inquiry.

(4) The Garvey problem and the New Dark Age problem are deeply connected, despite their differences. In both cases we find that the barriers to the rediscovery of recorded knowledge are sometimes higher than the barriers to fresh discovery. The barriers may arise from inadequate technology or cultural pathology. But either way, we find ourselves on the dark side of access barriers, either wanting or not wanting to tear them down.

Access barriers can be financial costs like prices, or psychic costs like the courage to seek out evidence and arguments that may contradict hallowed certitudes. Inquiry can be thwarted by our inability to gain access to relevant literature or by our inability to evaluate the literature after we do have access. When we're waking up to the opportunities of a vast new wilderness of knowledge and knowledge claims, we face the Meno problem in this form: sometimes we need a reliable guide in order to find a reliable guide.

Digitization, search, and OA are making steady progress toward banishing the classic Garvey problem. But they're not making the same steady progress toward postponing a New Dark Age. Our access system may be improving rapidly, but if a cultural whirlpool made serious research harder than submitting to our peers and elders, then this new dark Garvey variant would creep back like jungle over cropland.

(Note that right up to his death, [Ray Bradbury insisted](#) that the problem depicted in *Fahrenheit 451* was not top-down censorship and book burning but bottom-up tribalism and loss of interest.)

One obvious lesson is not to let access become harder than fresh discovery. That's an argument for OA, even if OA is far from sufficient to spread knowledge to those who are more or less willfully neglecting it.

Another lesson, which I hope is equally obvious, is not to be content with setting the bar that low. Eliminating Garvey problems is the least, not the most, that we can ask from an effective access system. Looking up the results of a past inquiry can be unconscionably costly and still less costly than redoing the original inquiry. The OA movement is right to try to solve the Garvey problem and right to keep pushing to make access easier and easier.

But even making access as easy as contemporary technology allows sets the bar too low. Even when looking up what we don't know is essentially costless, we must create a culture in which we want to do so. Until then we won't have answered Mark Twain. If we don't actually look up or learn what we don't know and need to know, and don't actually incorporate it into our lives, then we're no better off than people thwarted by Garvey problems who can't look things up.

There are good reasons why the OA movement focuses narrowly on price and permission barriers, not more broadly on the insidious barriers of will. But the similarities are real and worth bringing out for examination. When we do bring them out, and hugely broaden the problem of access barriers, we see that OA is necessary and sufficient for solving the narrower problem and necessary but insufficient for solving the broader problem. That's as optimistic as I can get.

Postscript. If it's elitist to prefer knowledge and research to ignorance, then I plead guilty to elitism. But it would be perverse to describe any position as elitist that wants to tear down access barriers and share knowledge and research with everyone. If you think that favoring knowledge over ignorance invites invidious or question-begging ways of drawing that distinction, I agree. I haven't forgotten that problem. I address it here in Part 2 in the form of the Meno problem. But I remind the reader of Part 2 that I also addressed it in Part 1 by framing the goal as access to "research" in the wider sense, not just "knowledge" in the narrower sense. ("We want access to all the data, evidence, [and] arguments...that help us decide what to call 'knowledge', not just to the results that we agree to call 'knowledge'. If access depended on the *outcome* of debate and inquiry, then access could not *contribute* to debate and inquiry. We don't have a good name for this category larger than knowledge, but here I'll just call it 'research'....") Hence, postmodern suspicion of "knowledge" does nothing to detoxify the prospect of a

New Dark Age.

“Discovery, rediscovery, and open access” © 2010, 2025, by [Peter Suber](#) is licensed under [CC BY 4.0](#).