10
Too Much Content?

Compared to other ancient civilizations, Athens had a high literacy rate—most well-born males in the city, tapering off rapidly elsewhere and among other groups.[1] But even for those who could read, there was not much to read.[2] How different our predicament!

We produce ever more content. Material to read is no longer our lack, the time to do so is. The real expense of content is not its price but the opportunity cost of what we could be doing otherwise.[3] "Science that is not seen does not exist" is the motto of RedALyC, a large Latin American open-access portal.[4] Being published is not the same as being seen, much less read. We now fight more about attention than publication. Many lament the surfeit of content, wishing to return to an allegedly simpler and sparser past. Few consider that data have grown faster than content. That suggests a need for more, not fewer, articles.[5]

What does too much content mean? Do the oceans have too much water? The question is nonsensical except within some constraint—for fishing, shipping, recreation, CO_2 absorption, or whatever. If every reader is expected to be aware of—never mind reading—each word of the growing mass of content, then we have an impassable dilemma. But if there is some division of labor, and

not everyone reads everything, then things look different. With a global production of 2.2 million books and 3 million articles, we put out some 200 billion words annually.[6] They clamor for attention from some five billion potentially reading pairs of eyes.[7] Ignoring newspapers, blogs, magazines, and the like, some forty new words annually await each reader.

That seems surmountable. Yet, averages are misleading—no one consumes one. Like lemmings, we cluster. In the US, the ten bestselling books in 2019 were bought 12 million times in print editions.[8] Though they represented but 1/30,000th of all published volumes that year, they accounted for 1/50th of total print book sales of 640 million units.[9] If readers' attention—not just buyers' discretionary dollars—was similarly concentrated, we would have 100 million readers globally poring over 73 books in any given year. Such focus on rare works reveals the power-law distribution of attention—that much content will be read by only a few, most by none. The long tail grows ever longer the more content we have. Finding and selecting what to read are as important as actually consuming it.

Keep in mind that words multiply, but so do readers. Researchers and writers are also readers, so an expansion of the audience is baked into an increase in authorship. As education widens and improves, readership mounts, too. The interested public is larger now than ever. How broadly readership is corralled varies among fields. In theoretical physics, authors and readers are largely the same people. But readers outnumber authors where practitioners and researchers overlap only tangentially, such as in medicine or engineering.[10] A sampling of articles in *Pediatrics*, for example, was read on average 14,700 times over their lifetimes.[11] Open access itself promises to attract a larger audience. It stands to reason that, as barriers fall, more consumers will be tempted to sample the wares. Journals that open up enjoy a bounce in readership.[12] Still, increased downloads do not always translate into more citations.[13] And in any case, the advantage is less than might have been expected if suddenly the candy store flung open its doors.

Open articles are more often cited than those locked behind paywalls, although, surprisingly, the advantage is modest, some 7% or 8% more. This reading uplift may be thanks to earlier publication than for printed versions. Moreover, the benefit may be greatest for the oldest and most prominent journals. In contrast, lesser-known periodicals suffer from the increased competition for attention allowed by open access, their viewings declining as they open up.[14] Sales of paper versions of digitized open books have also increased between 5% and 8%.[15] A study of Swiss monographs found no effect on sales but a large upturn in views.[16] A Dutch study revealed no effect on either buys or citations but increased online usage.[17]

Meanwhile, another analysis has registered whopping increases—seven times the downloads, ten times more mentions, and 50% more citations. But this was an in-house investigation by Springer of its own open books compared to its non-open works, so possibly it was not entirely impartial.[18] It stands to reason that the long tail of obscure content would—once digitized—benefit most from being discoverable. As measured over the twentieth century in all fields except the humanities, ever fewer articles go wholly uncited, and citations are dispersed over a larger group of them.[19]

Readers are not aimless. They seek what interests them, and the task is guiding them to pertinent material. The content avalanche is partly channelled as some fields organize their best output in a few core journals located at the pinnacle of attention. In less centralized subjects, such as history, that effect is weaker. Here, attention is fragmented into hundreds of national, temporal, methodological, and other subfields, each issuing work across a welter of outlets. But such fields are not those suffering from an overpowering flow of content.

Information abundance is also tamed by ongoing distillation. Textbooks and other secondary works mediate between the coalface and consumers. Once-canonical texts fade to become part of the intellectual background noise. In fields that study physical reality directly, the upsides of intermediation win outright. What does a medical student gain by reading outdated physiologists? Great

books courses in the humanities, in contrast, assume that renewed contact with classic formative texts repays the effort. Disputes here are over which works to include and how to limit an expanding canon to something manageable.

Without necessarily having read Plato, Marx, Weber, or Freud, most university graduates have a vague sense of what they argued, thanks to endless summaries and recapitulations in secondary works. Sometime in the future, everyone other than a few hardcore originalists will agree that the caravan has moved on. Even figures of this stature will have become of primarily historical interest. In the long run, even the greatest thinkers live on only in their distillation.

Too much content? If so, in what sense? From the earliest libraries, readers have complained of too much. Seneca the Younger, Stoic philosopher of the first century CE, condemned large libraries. What was the point of having so many books that no one could read even their titles in a lifetime?[20] Printing worsened matters. In 1525, Erasmus complained of a wave of new books that distracted attention from worthier works of antiquity.[21] In 1996, Umberto Eco echoed the lament, presumably tongue firmly in cheek. There were too many books, and he was hoping for some relief as would-be authors instead channeled their efforts through the then-novel medium of e-mail.[22] Fat chance. Books have continued to proliferate, e-mails even more. And no one would ever complain about a surfeit of content who had ever set foot in a bookstore in the old East Bloc, with its selection of Marxist-Leninist classics, manly proletarian novels, and paeons to overfulfilled five-year plans.

Storage and Memory

With digitality, the cost of storage diminishes dramatically. Already 25 years ago—eons in digital lifespans—we were on the cusp of sufficient capacity to save everything—every printed page or writing,

film, photo, TV and radio program, music recording, and phone call.[23] Since then, social media and the web have added massively to content, but storage has expanded even more.

Kryder's law is the correlate to Moore's on the increasing power of semiconductors, and predicts a continuing exponential drop in storage costs.[24] The measure at the turn of the century was terabytes. Today, 20 years later, it is zettabytes, a trillion times larger. The equivalents of digital storage devices that those of a certain age remember paying hundreds of dollars for—whirring, blinking, chunky, crashing bits of kit—are now given away as advertising gimmicks on keychains.

The information we need is probably already out there, possibly multiple times, but if we cannot locate it, it might as well not exist. It is silly, however, to reinvent wheels that merely need to be found. That strikes social scientists as a truism. Yet, for natural scientists, who study the same reality as past colleagues, it may be more efficient to probe nature anew than to determine whether anyone has previously made similar observations. Descartes put his finger on the problem. Books mixed knowledge haphazardly with dross. Better, therefore, to go straight to the coalface, mining reality oneself.[25] But for scholars who regard even past errors as interesting, scrubbing away inherited accretions is profoundly wrong, even in quest of renewal and cultural rebirth.

As more information is stored, duplicative innovation becomes likelier. Yet, with each iteration salted away and with search tools available, subsequent inventors, discoverers, or formulators should be better able to find antecedents before committing redundancy. We can glimpse a world where, if someone somewhere has ever known something, the rest of us can as well, and wheels are never reinvented.

Data surfeit also raises the problem of misinformation. More data are available, but much is bunk. The problem is, as the old joke about advertising goes, it is unclear which part is dreck. No one who

has read the readers' comments at the bottom of online articles in even reputable newspapers ever regains their full faith in humanity. The world is full of creeps, jerks, fantasists, knuckle-draggers, the delusional, the very, very angry, and the just plain stupid. For those high-functioning enough to both surf and type, the internet has provided a forum for the first time in human history. The global id is not a pretty sight.

We have touched on the cultural consequences of vast available content and the danger of future creators being discouraged or seeing their role as reusing existing materials rather than creating *ab ovo*. Does too much information undermine our ability to act, vitiating decisiveness in the face of innumerable demands on our attention? Are we becoming a culture of archivists?

Memory and storage are distinct. One affects our psyches directly and unavoidably. The other—external to us—can be tapped as we wish, without obligation. Our collective store of cultural knowledge has been accumulating from before humans invented writing. Oral traditions were no more durable than their carriers. We know of them only insofar as they were committed to some more permanent medium, whether by the scribes who immortalized Homer or the brothers Grimm for Germanic fairy tales. Writing massively expanded our collective cultural storage, as did printing. Not until the digital did we again take such a radical leap. The in-between technologies, such as microfilm and microfiche, were but minor expansions, although enthusiasts at the time thought they might make entire libraries available to the world.[26]

A transactive memory system has multiple people recall information collectively, distributing responsibilities for different aspects. The internet has become our partner in such respects. Psychologists distinguish a feeling-of-knowing something that is actively in our memories from a feeling-of-findability, a sense of how readily we can locate information.[27] As we increasingly rely on the internet as surrogate memory, we become part of a larger intermind.[28]

Whether that is memory or storage need not detain us. Recall Michael McCormick's account of finding evidence of Arab coins in medieval Europe. His hunch that the information was somewhere in Widener paid off. And he lauded previous librarians' foresight in supplying subscriptions to arcane Belgian numismatic journals. That was a hymn of praise to old-fashioned libraries, uninformed by the web's possibilities. Today, with Google Books largely online, we enjoy the assurance that information is out there and must only be located.

Storage presents a problem only in the physical world when we run out of space. Few garages are actually used to shelter cars from the elements. Along with attics and cellars, they store stuff. Only when we reach capacity is there an issue. In digitality, however, space is no longer scarce. Attics have become infinitely expandable. We are hardly conscious of what is stored or even that it is being kept. We live our lives psychologically unencumbered by considerations that, as a society, we have become data packrats. We do not seem paralyzed by too much information. Indeed, one could argue that the more we store, the less we have to remember, the less we are weighted down.

Socrates railed against writing as vitiating true culture. In the *Phaedrus*, he argued that writing will atrophy humans' memories. Relying on writing—marks made by others, external to ourselves—will banish learning from our minds. Plato thought that knowledge from mere reading without contact with a teacher was not true understanding.[29] And indeed, psychologists have documented an offloading effect of computer memory and the internet. Information known to be stored is more likely to be forgotten.[30] And rightly so! What is the point of externalizing data, whether writing or uploading, if not to free our minds for other tasks?

Nietzsche may have been correct that too much memory immobilizes us. But storage spares us the effort of recalling trivialities. Who today remembers phone numbers, all nicely stored on our devices,

with our intimates on speed dial? How long has it been since most adults performed long division without benefit of a calculator? Evolutionarily we have outsourced much of our digestive function to fermentation and fire, processes external to ourselves that allow us to absorb nutrients higher up the food chain. And we share such activities with the gut biome we acquire as a symbiotic helpmate to tackle the world more efficiently. Without such shortcuts, humans would be like ruminants, spending our days laboriously digesting plants. Philosophers now discuss distributed cognition as a form of knowing—not just remembering—performed by collectivities that are otherwise bereft of the psychological unity of individuals that normally explains how humans understand.[31]

Writing expanded our mental range. It allowed us to outsource what we now regard as the triviality of committing to memory. Culturally, we became dependent on parchment and paper. Despite some destructions and book burnings, written documents were more capacious and robust than the memories of bards and minstrels.

Digital storage takes us further along a road we have been traveling for centuries. Much is automatically backed up. Every revision and addition to our writings is retained for those who care to know. We leave wide data trails, allowing our lives to be uncovered retrospectively. This information surfeit is not much of an issue for us, except when we are defending our privacy. It is more one for future historians, faced with an embarrassment of riches and needing to sort the excess. As problems go, there are worse.

Authors vs. Readers

Mega-journals, online repositories, and other gushing founts of little-curated, underreviewed, overprovisioned content provoke unease. Letting down the guard of publisher review has created a

supply problem. Vast pools of content magma bubble below the surface. Given an outlet to erupt via digital repositories and other low-barrier media, volcanoes of raw content now spew forth. Their deluge washes out the limpid rivulets of publisher-curated content with a vast torrent of good, bad, and indifferent verbiage. A flood of junk science and other would-be knowledge inundates us.[32] As with money, bad knowledge drives out the good. That seems to be the fear.

In this scenario, authors have become the reader's enemy. Best-selling, well-paid writers, the beneficiaries of their publishers' careful curation and promotion, may suffer from a muddying of the waters as new content rushes in to compete for the audience's strained attention spans. Most self-published books are in genre fiction. Whether that detracts from established authors in these niches is unclear.

For academic authors, however, the incentives are different. High-volume, low-selection publishing on the model of *PLOS One* and repositories such as arXiv have been criticized as good for authors but bad for readers.[33] Publishers serve readers—so goes the assumption here—by filtering out low-quality work that would otherwise distract the oversupplied consumer from the best.[34]

Academic authors are more interested in publishing than selling their work. In academe, sales are unimportant. They can be measured only with books and are irrelevant for journal articles. What matters in academia are impact, audience, and citation. Most readers of scholarly literature borrow, not buy. If hiring, promotion, and funding depend on not just publications' quality but also their quantity, some scholars may publish as much as possible. New means of dissemination that issue material without the barriers of peer review may tempt them to empty their desk drawers and fill the repositories.

Online, low-obstacle publishing may reduce the pressure to self-select and self-curate content, sinking authors' shame thresholds.

In the current system, only the need to await results discourages authors from submitting their manuscripts to the most prestigious journals. That burdens these periodicals. *Nature* receives over ten thousand manuscripts annually.[35] In effect, they ration access by time, as manuscripts await review. Money could be added via submission fees paid in addition to publication charges. Authors would be encouraged to think twice before submitting, and selective journals with the heaviest loads would be compensated, removing one argument for high publishing charges.[36]

Compared to the old system of review and filtering, does a larger percentage of all content today see the light of publication than earlier? Some argue that digitality and self-publication allow almost everything ever written to emerge. The total number of published works is therefore increasing.[37] That holds for the self-published books that are merely reproductions of public-domain material. New volumes may issue, but not new titles. But this logic of current surfeit works only on the assumption that in the old system, a rejected manuscript was withdrawn from circulation altogether, never to see the light. That is far from clear.

Although particular publishers and journals may have been selective, the system as a whole was not. If works were resubmitted until they found an outlet, individual selectivity was compatible with overall ecumenicity. In the legacy system, publishers were arrayed along a cascade of prestige. Works rejected by one journal or publisher were usually resubmitted to another, eventually finding their resting place in the hierarchy.

A stream of submissions used to inundate publishers, with rejected manuscripts receiving multiple reviews until they finally came to rest. Did a larger percentage of manuscripts than now remain unpublished, even after running the gauntlet? What happened to manuscripts submitted to any given publisher? Was there market clearance of the slush pile? No single publisher could know, but studies often take them at their word, that rejected manuscripts

just vanished back into authors' drawers.[38] Publishers' assurance that their particular rejection settled matters once and for all speaks to their vanity.

With conventional journals, was selection so rigorous that many manuscripts never appeared in any venue? Without a universal register of submissions to which the published outcome can be compared, we cannot know. One study in 2001 reported that slightly more than half of initially rejected manuscripts were ultimately published elsewhere. But this examined editors who had looked to see whether manuscripts they had spurned were eventually issued, therefore likely to be an underestimate.[39] Of manuscripts not accepted by the *Journal of Clinical Investigation*, 85% later emerged elsewhere.[40] Similar statistics can be multiplied at will.[41] One observer was only slightly exaggerating in his conclusion that "if a paper is submitted once, it will ultimately be published, some day in some journal."[42]

Of manuscripts in bioRxiv, a prepublication repository intended as a forum to improve works before formal submission, a third remains unprinted in conventional journals.[43] In arXiv, the physics preprint repository, 64% of articles posted also appeared in journals indexed in Web of Science. This varied from a high of 80% in condensed-matter physics to a low of 20% in computer science, where conference proceedings and their posting are the preferred means of information exchange—in other words, where formal publication is not the aim.[44] Only a small fraction of content, then, is likely to remain in the author's bottom drawer. Has this changed in the era of mega-journals, repositories, and predatory periodicals? Would it further change with a global bulletin board? Even if every manuscript, however often rejected, was posted in the ether, it seems unlikely to increase the amount of disseminated content dramatically.

We also know that much scientific work results in no publication at all. Over 40% of research work in medicine is not issued

as articles.[45] That is independent of digitality or open access. Nor is there reason to think that more formerly unpublished work now appears in open repositories. But it does raise the question of whether, with lowered barriers to dissemination, a backlog of work would now emerge in print. Negative results spurned by conventional journals with space constraints might be more likely to appear. Most of the "grey literature" of government studies, reports from nonprofits, preliminary research results, project websites, data archives, and the like were previously never formally published, but could now find an appropriate lodging in online repositories.[46]

Some university systems consider, or even emphasize, the quantity of publication. In China, the number of papers indexed in certain databases is often the basis of advancement.[47] If hiring, promotion, and funding rest with bureaucrats unversed in the fields they administer, quantitative metrics paper over their inability to make qualitative distinctions.[48] Such incentives can be perverse. Indisputably, rubbish has been published. Evidence from predatory journals in China indicates that articles can be ghost-written and plagiarized, and sometimes they are mashups of others' works.[49] Authorship and whole papers can be bought.[50] Since scientists in China earn more than doctors and lawyers, and rewards for publishing in top journals reach deep into six figures, the pressures are intense.[51] In such a system, more sophisticated quantifiable metrics would be an improvement. The h-index, for example, allows a more nuanced view of quality than just the number of works published.[52]

Nor does China stand alone in such respects. Melbourne Business School pays faculty bonuses of $A15,000 for articles published in one of the *Financial Times*' roster of quality journals.[53] When Australia started promoting scholars according to publication quantity, they issued more but worse papers.[54] In Serbia, it was considered a desirable reform of long-entrenched cronyism in academic advancement to require a specified number of articles in impact-factored journals for promotion instead. Alas, the new system was gamed by

periodicals that convinced Thompson Reuters to index them while still publishing whatever dreck came over the transom, so long as publication fees were paid.[55]

In the Czech Republic, a preternaturally productive young scholar, Wadim Strielkowski, gamed the system with 60 articles and 17 monographs, all issued by dubious publishers over a quick three years.[56] South Africa's government pays universities a handsome subsidy for every article published in journals indexed in SSCI or SCI.[57] Once a 2010 law required them to meet productivity thresholds for promotion, academics in Italy began citing their own work assiduously to inflate impact factors.[58] Yet, that rule had been a well-intentioned attempt to combat cronyism and nepotism. In well-regarded Western universities, students in three-year life sciences doctoral programs are expected to publish two first-authored articles. Much salami-slicing of research projects is motivated by such requirements. A dean of biological sciences at a UK university detailed how faculty candidates had been chosen solely for the number of their publications, the quality of the journals, their h-index, and other purely quantitative indicators. After reforms, candidates submitted three articles that were actually read by the hiring committee.[59]

Such misincentives are stronger when there are fewer barriers to publication. But ultimately, less-restrictive dissemination is how the problem is expressed, not its source. The incentives arise from universities rewarding quantity. Unless thus prompted, why would anyone, except incurable graphomaniacs, issue their merest scribblings? Arguably, the open-access fora, even the mega-journals with their cheaper fees, impose a greater barrier to the free flow of bilge than subscription periodicals. Conventional outlets require review, but as we have seen with the *Social Text* scandal and others, that is not always an obstacle.

With gold access, at least the perpetrator of arrant nonsense has to pay. Indeed, one could imagine a reverse auction where

mega-journals imposed some review and charged fees that rose as the apparent quality of submissions declined. It would be much like the marriage market in simpler times, which was cleared by offering higher dowries for women who found fewer takers.[60] Or like scholarships for clever students, with full-freight tuition for the less so. Admitted on merit, scholarship students at Oxbridge once wore special gowns proclaiming their distinction. If the variable fees charged for each article were made known, this would create in one fell swoop a publicized ranking system, an economic disincentive for unloading sub-par content, and a means of making mediocre but insistent authors subsidize costs.

To expect the dissemination system to throttle the motor of hyperpublication seems misguided. Nothing in digitality or open access requires universities to abandon their own criteria of merit. Even if they do not actually read and evaluate their researchers' publications and instead rely on postpublication metrics, such indicators—used sensibly—should separate wheat and chaff.

Overpublication?

Critics of open access often assume that scholarly careers are driven by quantity as much as quality.[61] Perhaps there is a grain of truth to this in some systems, such as the Chinese, that aspire to a larger presence on the global stage, have not yet arrived at maturity, and cannot be taken as characteristic. And possibly it is an issue in prolific fields where no one can reasonably keep up with their colleagues' output. The quantitative metrics mean more in hyperactive areas. Whether the articles whose quality the metrics supposedly measure are subscription or open access is irrelevant. But they allow hiring, promotion, and funding decisions without reading the material.

Yet, there persists a broader assumption behind such worries about excess content. Some think there are simply too many publications.[62] Others see a more subtle but equally insidious variant of the reward for quantity we have found in some institutions. This is attributed to a neoliberal quantification of intellectual metrics, a Fordist emphasis on measurable output. Neoliberal university systems pressure scholars to publish more than earlier and more than they want to.[63] The implication is that the aim is not better and more quality but has spun off into a cycle of pointless publication.[64]

A nostalgia for a vanished world of gentlemanly leisure pervades such accounts of our current malaise. It is much like the sepia-tinted view of past parenting practices. Cultural nostalgists often long for the early postwar era when intact nuclear families allowed children to be raised by loving stay-at-home mothers. Instead, we now have harried two-career couples, ordering take-out as they quarrel over whose turn it is to pick up junior from daycare. In fact, the unjaundiced eye of social science surveys reveals that parents spend more time today with their offspring, despite the prevalence of working mothers and heavier workloads for all.[65] In the good old days, mothers skimped on cookie baking to play tennis. Today, we work harder at parenting as well as our jobs.

We have less time for hobbies (inane enthusiasms that once passed the idle hours) and snoozing on Saturday mornings. Perhaps that is modernity's curse. But it is no more obviously a decline than is the modern academic's productivity compared to the patrician leisure, the life of reading and doing nothing with the knowledge gained other than teaching a few entitled undergraduates, which used to characterize university faculty. Sure, young academics at top universities have to publish more than their forebearers.[66] That is what upping the game means. Undergraduates, too, have to amass better records to get into these institutions. Bankers also work harder than their two-martini lunch predecessors. Nor has medicine

been resting on its laurels. Name a profession where harder work is not the norm. Increased productivity hardly seems worth a remark, much less complaint. Lest we attribute this solely to capitalism and neoliberalism, keep firmly in mind that Soviet five-year plans—innocent of market forces—were relentlessly overfulfilled by hard-working Stakhanovites.

Do scholars overpublish? In the humanities and social sciences, the issue simply does not exist. Scholars publish a book at most every several years, the vast majority never more than one, max two, over their careers. Insofar as they write articles at all, perhaps one every year or so. The main fault here is that humanities articles are sometimes derivative of books, duplicating information scheduled to appear as chapters. The hard sciences and some of the social sciences, in contrast, may have a problem. We have touched on hypertrophied authorship, both the pile-on of multiple authors in the scrum and individuals' preternatural prolificacy. Scientists may issue more publishable units than before. Some observers claim that they achieve this by morcellizing output, salami-slicing projects into more and shorter pieces.[67]

Insofar as it occurs, salami-slicing research into multiple publications suits personnel decisions that reward quantity and the publishers who issue it. But it increases the work of reviewers. Not only do they have more articles to read, but they must also consider the entire nimbus of manuscripts surrounding the one in question to rule out unwarranted duplication.[68] Unfortunately for peer reviewers and librarians who shoulder the burdens of hyperpublication, multiple articles generate higher overall attention metrics. Citations increase when projects are published in many papers, not least because authors cannot resist the temptation to cite themselves.[69] Unfortunately for readers, the sweet spot for maximizing citations appears to be both many and larger articles.[70]

Some observers have suggested limiting how many works should be permitted researchers annually. Following the slow food

movement's lead, they think that reflective, snail's-pace scholarship would enhance quality.[71] And some funders have sought to calm the supposedly roiled waters of publication by restricting the number of articles included on applications.

Debates on overpublication must be conducted within each discipline. Outsiders cannot judge whether prolixity is a problem. Publication differs widely among fields. Books play no role in some, articles little in others. And articles serve different functions. At one end, we have statements of much thought and work in fields such as philosophy.[72] Decades hence, the best will still be read. Conversely, in the life sciences and medicine, articles are often just short reports to establish priority and keep colleagues abreast of results. These outpourings may seem overwhelming to philosophers, but each piece is correspondingly ephemeral, not intended to be read even a few months hence as new results emerge and the caravan moves on.

Some fields internalize the review process, taking time to present ideas at talks and seminars, soliciting feedback before publication. Others externalize it, publishing quickly and awaiting responses. We would not want to confuse speed or quantity with quality. But neither should we assume that any increase in publication tempo or output necessarily indicates a decrease in worth. Sometimes more is more.

The output of academic content has increased in line with the growing number of scholars globally. That is not the result of open access and would have held in the old subscription system as well. The effect of mega-journals and online repositories will hit readers less than it will affect other, existing journals. Readers will have to grapple with the overall increase in content however it is delivered. But the immediate consequence will be to heighten competition at the lesser ranks of the publishing food chain. *PLOS One* and other mega-journals do not compete with top-ranked periodicals. Instead, they compete for content with lower-placed journals, which offer

less as substitutes for the new outlets. That may portend a shake-out in the industry. Still, it leaves the average reader no worse off and arguably better served by the new open alternatives to yesterday's lesser subscription journals.

On the other hand, we have a different problem if mega-journals and repositories are encouraging more and worse work that earlier would not have been written, would not have been submitted, or would have been rejected by even the least-discerning conventional journal. Is output growing per capita, more for each of the increasing numbers of scholars? Only then would it be true that mega-journals and open access have diluted the quality of published research. Only then could one plausibly argue that open access encourages a downward trend whereby "every paper authored would be published, regardless of its quality."[73] How true is this compared to subscription journals?

More to the point, even if more content per capita is published today, even if some is mediocre, and even if that would formerly never have seen the light—is this necessarily a bad outcome? There is no book so bad, said Pliny the Younger's uncle, that some good cannot be wrung from it.[74] Many assumptions behind these questions need unpacking. Is there a large backlog of unpublished material that can now be issued in the repositories? If so, is this material unpublished because it is mediocre? Or because the established publication channels are swamped and unable to process it? Because the authors cannot afford publication charges or surmount other obstacles? Because established subscription journals apply evaluation criteria that exclude scholars from certain institutions, backgrounds, or countries? Because other nonscholarly obstacles hinder some researchers from issuing their findings?

We need to know two things. Compared to the era of conventional publishing and subscription journals, are there extra publications per capita that see the light with a switch to low-barrier dissemination? Or is the undeniable increase in academic publishing of

recent years due to new authors entering the field as universities expand and once-excluded nations join the global research endeavor? If the former, then open access may be a factor. If the latter, not so much. The growth of worldwide research is a welcome development and, in any event, is not due to easier publication.

Globally, there were six million researchers a decade ago, more recently between seven and eight million. Of these, about 20% are repeat authors.[75] These numbers have grown much. In the US, engineers and scientists expanded tenfold from 0.26% of the labor force early in the last century to 2.5% in 1970. Mathematicians and information technologists increased 500-fold over half a century. In 2008, knowledge-based professionals made up 20% to 30% of the labor force in developed nations.[76] More researchers mean more research. China's output of scientific articles has increased 20-fold, from 6,000 in 1990 to 123,000 in 2011. By 2011, Chinese scientists published two-thirds as much as their US colleagues, including only articles indexed by Thomson Reuters. Between 2008 and 2014, the Chinese share of all scientific articles doubled from 10% to 20%.[77] China produces the second largest number of papers, after the US. But scientists (including engineers) make up 0.4% of the labor force in China, compared to 3.1% in the US.[78] The Chinese supply of content thus has ample headroom to expand before it justifies suspicions that more research is diluting standards.

Compared to half a century ago, most of the enhanced amount of content produced today comes from more researchers, not more output per head. That content is increasingly salami-sliced, subdivided into more and smaller articles, is something of a myth. More work is today coauthored. Dividing papers by the growing ranks of coauthors reveals a stable or even declining real output per head.[79] Other studies confirm that per capita research publication has been broadly constant over the past century.[80]

But let the devil have his say. Would it matter if we had proportionally more content today and each researcher authored more

than earlier colleagues? Take the question a step further. What if this surplus content is also more mediocre than in the past. Why is that a problem? Once we have more works than we can ever hope to read, selection becomes a necessity. As mentioned, Seneca the Younger in the first century CE complained about that. And once sorting is required, it matters little how much is classified as dross for any given act of choosing. We already have more than we can read. *How* much more is irrelevant.

The per capita output of books ranges widely among nations. Even the measures are sometimes wonky. The British pride themselves on having the highest rates, but that is a statistical artifact.[81] UK and US publishers have long agreed on a monopolistic division of the market.[82] Books in English were issued exclusively in their respective spheres by UK and US presses. The Americans got the Philippines and sometimes Canada, the British, some 70 nations, once from the Empire and many later from the Commonwealth. UK books-in-print numbers are inflated by counting the UK editions of US books. If the Austrians required every German book sold there to be issued in a local edition, as the UK does for US books and vice versa, Austria would be the per capita publishing powerhouse of the world. Something similar must hold for Canada, not to mention Australia and New Zealand.

Iceland publishes more books per inhabitant than anyone. That is unsurprising for so small a country (population 357,000) with its own language. Even if it issued only a phone book, dictionary, thesaurus, and an encyclopedia, it would be ahead of the game. But, in fact, Iceland sports a vibrant publishing industry, all the more impressive for its dollhouse size. Per capita, Iceland publishes 238 times as many books as Kenya. Even compared to the Germans—bibliophiles who invented the medium—it issues two and a half times as many books, and almost four times compared to the Swedes.[83] So are the Icelanders overprovided with content, or are the Swedes deprived?

Where is the downside to as much information as possible, even if it is not all equally good? A similar discussion breaks out every time university admissions are expanded. Only a tiny fraction of 20-year-olds attended tertiary education in the old elitist system; today, one-third of the cohort does. The nature of universities has changed accordingly. Colleges can take less preparation for granted; they have had to broaden standards and to remediate. But besides a few gnarly classics dons, would anyone want to return to the old system?

Granted, we are becoming more credentialized. The BA that in 1890 would have guaranteed a job at the State Department is now an entry-level qualification to steaming lattés at Starbucks. But what of the advantages of having flight attendants noddingly acquainted with the concept of orientalism or IT support staff who can distinguish a mean from a median? Lucky the society able to allow so many such education! Much the same holds for content superfluity. Even if the average BA today no longer reads Latin, the overall cohort of 20-year-olds is better trained than a century ago. Even if not every work is comparable to yesterday's best, and even if some is shoddy, the quality, quantity, and usefulness of content today are overall better.

MA theses are generally neither published nor even archived anywhere. They end up tucked into corners of authors' attics or hard disks. If they were uploaded to depositories instead, the bulk of the world's knowledge would not change. Though currently largely invisible, MA theses are out there, and, once uploaded, they would become more useful. Of course, not every MA thesis is worth wide distribution or readership, but some are. The world would not be worse off if MA theses were suddenly searchable and readable, any more than it is with information overload, even if it is not all equally good. As the saying goes, there is no such thing as bad weather, only unsuitable clothing. There is no such thing as too much information, only inadequate search engines.

Indisputably, we have more information than ever. Is it too much? Even serious studies, not just grumpy old dons, complain about "overly abundant scholarly information."[84] What does too much mean? As the selection tools improve, surfeit should fade as an issue. Sophisticated search engines will tame the content we choose from, allowing us to focus. In any case, it is churlish to complain of excess. The question is, would the world be better off with less?

Economists and psychologists point out that too much choice can provoke anxiety, confusion, and bad decisions.[85] *Free to Choose*, Milton and Rose Friedman's paean to individual liberty, became *Free to Lose*, John Roemer's cautionary tale on the perils of unbridled markets. That is psychology, not epistemology. It holds for the supermarket, not the hardware store. Perhaps we needlessly fret over what to put in our mouths, anoint our skin with, or wash our hair in. But who believes it is better to have fewer tools to repair with, fewer medicines to treat with, fewer concepts to analyze with? Only rarely is surfeit an annoyance. And the obvious bears restating: mediocre, misleading, and mendacious stuff already bloats the web. Enormous quantities of bilge wash through the internet's portals. Open access, in contrast, allows a presence also for the good, the scholarly, the footnoted, the researched, and the reviewed. Again, sometimes more is indeed more.

Even the minimally vetted content posted in repositories, awaiting its readership and critics after the fact, comes with guarantees of quality that the web's bloviations lack. First, even mediocre scientific articles serve a purpose. The long tail is lengthy indeed. Who is to say that not everything finds its spot? Since dissemination costs little, nothing is lost by launching everything and seeing what arrives onshore. Because authors pay for gold access, arrant nonsense must surmount a built-in barrier. Who would front fees to float something they did not think was worth the electrons that fire the pixels that make it legible—or whatever is the digital version of "the paper it is printed on"? The idea that scholars busy themselves

churning out garbage to pad their CVs, paying to clog the ether and hog attention, is wildly exaggerated.

We are afraid of being buried in junk science. But where would it come from? Who are these people with the time and funds to run labs producing bad research? You can do it once, but then your reputation suffers. Even at the dawn of digital distribution, Paul Ginsparg, founder of arXiv, argued that the very act of distributing content widely encourages authors to self-regulate. Researchers are keen to ensure that their posted content not embarrass them before a now-magnified audience.[86]

Yes, digitality has unleashed content. Self-published books supply an avalanche of new material that would not otherwise have seen the light. Most are fiction of various genres, hard to compare with academic work. Scholars write to probe and reveal the truth but also to attract notice and advance careers. For each of these motives—especially pursuing truth—they must toe certain lines. Regardless of how they publish, academics write within a framework of evaluation they have internalized, whose criteria they work toward. Unless their work meets scholarly standards, it does them no good.

"Without peer review, we are nothing but well-paid bloggers," one cynical observer of his fellow academics notes.[87] That is fundamentally wrong. Authors who do not follow their discipline's epistemological precepts even when not policed by formal peer review are not scholars. Nor would their work enjoy renown, or their careers receive advancement. In the broader sense of judgment leveled by colleagues, review is omnipresent in academia's panopticon, not just at the instance of passing through the needle's eye of formal evaluation for publication.

In the days of prepublication review, presses were not swamped with hopeless submissions. That books full of rubbish were only rarely published probably says as much about the quality of the intake as about any heroic acts of selection by editors. It is unlikely

that only editorial evaluation spared readers inundation by printed bilge. The worst enemies editors face down are mediocrity, overspecialization, and tedious prose. Few authors would have bothered writing a submittable manuscript of nonsense. Why should that have changed?

Just because something is easy to bring out does not mean it is simple to produce. Yes, there are "books" for sale on Amazon consisting of cut-and-pasted Wikipedia content. No, the world is not improved by their presence. But these are like snake oil flogged to rubes, signs of an immature consumer market. The answer is not to forbid snakes or oil, but some combination of education and regulation. Eventually, such misuse will be relegated to the future's equivalent of magazine back pages or late-night TV ads.

Admittedly, plagiarism is easier to commit when only cut and paste are required, not even wielding pen and ink. Nor is plagiarism a crime, but merely a moral blemish. Moreover, certain aesthetic theories validate it as an element of derivative creativity.[88] But plagiarism is also easily exposed by the same technologies that facilitate it. Search engines tuned to such purposes routinely scrutinize undergraduate essays for tell-tale signs of duplication. Colleges hammer home to students that there is no excuse for committing plagiarism just because it can be done.[89]

It is a cultural curiosity that central Europe extravagantly valorizes the doctoral degree. In Germany, most universities are regarded as broadly equivalent in stature, whatever the reality. Ambitious politicians cannot burnish their CVs by attending the equivalent of Oxbridge or the Ivy League. Instead, they must scale the ranks, adding a postgraduate diploma. The outcome has been a rash of plagiarized doctoral dissertations in central Europe.[90] Entire websites are now devoted to scouring German politicians' dissertations for duplicity.[91]

In contrast, a doctoral degree is more a liability than an asset among aspiring politicians in the Anglosphere. US senators and

congresspeople sport professional degrees galore, but rarely PhDs.[92] With a doctorate in history, Newt Gingrich was a recent exception. Even less credentialled are politicians in Britain. Most MPs (90%) have no more than a BA, often in liberal arts subjects that would not obviously qualify them to run a modern industrialized economy.[93]

In the academic world, plagiarism is not a serious issue. Scholars are socialized into producing verifiable work. Their reputations depend on convincing their peers of its worth. Especially in the sciences, research and its results are costly and complicated. Long before the point of review, whether before or after publication, the system discourages malarkey. The suspicion that legions of scholars are churning out substandard or even fallacious work, hoping to flog it to an unsuspecting public, is bunk.

Athena Unbound

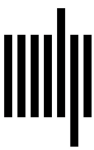
Why and How Scholarly Knowledge Should Be Free for All

By: Peter Baldwin

Citation:

Athena Unbound: Why and How Scholarly Knowledge Should Be Free for All
By: Peter Baldwin
DOI: 10.7551/mitpress/14887.001.0001
ISBN (electronic): 9780262373968
Publisher: The MIT Press
Published: 2023

The open access edition of this book was made possible by generous funding and support from the author



The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in ITC Stone Serif Std and ITC Stone Sans Std by New Best-set Typesetters Ltd.

Library of Congress Cataloging-in-Publication Data

Names: Baldwin, Peter, 1956– author.
Title: Athena unbound : why and how scholarly knowledge should be free for all / Peter Baldwin.
Description: Cambridge, Massachusetts : The MIT Press, [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022027103 (print) | LCCN 2022027104 (ebook) | ISBN 9780262048002 (hardcover) | ISBN 9780262373951 (epub) | ISBN 9780262373968 (pdf)
Subjects: LCSH: Open access publishing. | Scholarly electronic publishing.
Classification: LCC Z286.O63 B35 2023 (print) | LCC Z286.O63 (ebook) | DDC 070.5/7973—dc23/eng/20220628
LC record available at https://lccn.loc.gov/2022027103
LC ebook record available at https://lccn.loc.gov/2022027104

10 9 8 7 6 5 4 3 2 1