Why books don't work

Books are easy to take for granted. Not any specific book, I mean: the *form* of a book. Paper or pixels—it hardly matters. Words in lines on pages in chapters. And at least for non-fiction books, one implied assumption at the foundation: *people absorb knowledge by reading sentences*. This last idea so invisibly defines the medium that it's hard *not* to take for granted, which is a shame because, as we'll see, it's quite mistaken.

Picture some serious non-fiction tomes. *The Selfish Gene*; *Thinking, Fast and Slow*; *Guns, Germs, and Steel*; etc. Have you ever had a book like this—one you'd read—come up in conversation, only to discover that you'd absorbed what amounts to a few sentences? I'll be honest: it happens to me regularly. Often things go well at first. I'll feel I can sketch the basic claims, paint the surface; but when someone asks a basic probing question, the edifice instantly collapses.

Sometimes it's a memory issue: I simply can't recall the

relevant details. But just as often, as I grasp about, I'll realize *I had never really understood* the idea in question, though I'd certainly thought I understood when I read the book. Indeed, I'll realize that *I had barely noticed how little I'd absorbed* until that very moment.

I know I'm not alone here. When I share this observation with others—even others, like myself, who take learning seriously—it seems that everyone has had a similar experience. The conversation often feels confessional: there's some bashfulness, almost as if these lapses reveal some unusual character flaw. I don't think it's a character flaw, but whatever it is, it's certainly not unusual. In fact, I suspect this is the *default* experience for most readers. The situation only feels embarrassing because it's hard to see how common it is.

Now, the books I named aren't small investments. Each takes around 6–9 hours to read. Adult American college graduates read 24 minutes a day on average, so a typical reader might spend much of a month with one of these books. Millions of people have read each of these books, so that's tens of millions of hours spent. In exchange for all that time, how much knowledge was absorbed? How many people absorbed most of the knowledge the author intended to convey? Or even just what they intended to acquire? I suspect it's a small minority [1] Unfortunately, my literature reviews

have turned up no formal studies of this question, so I can only appeal to your intuition..

I'm not suggesting that all those hours were wasted. Many readers *enjoyed* reading those books. That's wonderful! Certainly most readers absorbed *something*, however ineffable: points of view, ways of thinking, norms, inspiration, and so on. Indeed, for many books (and in particular most fiction), these effects are the point.

This essay is not about that kind of book. It's about explanatory non-fiction like the books I mentioned above, which aim to convey detailed knowledge. Some people may have read *Thinking, Fast and Slow* for entertainment value, but in exchange for their tens of millions of collective hours, I suspect many readers—or maybe even most readers—expected to walk away with more. Why else would we feel so startled when we notice how little we've absorbed from something we've read?

All this suggests a peculiar conclusion: as a medium, *books are surprisingly bad at conveying knowledge, and readers mostly don't realize it.*

The conclusion is peculiar, in part, because books are shockingly powerful knowledge-carrying artifacts! In the *Cosmos* episode, "The Persistence of Memory," Carl Sagan exalts:

What an astonishing thing a book is. It's a flat object made from a tree with flexible parts on which are imprinted lots of funny dark squiggles. But one glance at it and you're inside the mind of another person, maybe somebody dead for thousands of years. Across the millennia, an author is speaking clearly and silently inside your head, directly to you. Writing is perhaps the greatest of human inventions, binding together people who never knew each other, citizens of distant epochs. Books break the shackles of time. A book is proof that humans are capable of working magic.

Indeed: books *are* magical! Human progress in the era of mass communication makes clear that some readers *really do* absorb deep knowledge from books, at least some of the time. So why do books seem to work for *some* people *sometimes?* Why does the medium fail when it fails?

In these brief notes, we'll explore why books so often don't work, and why they succeed when they do. [2]Let's get it out of the way: I'm aware of the irony here, using the written medium to critique the written medium! But if the ideas I describe here prove successful, then future notes on this subject won't have that problem. This note is mere kindling, and I'll be very happy if it's fully consumed by the blaze it

ignites. Armed with that understanding, we'll glimpse not

only how we might improve books as a medium, but also how we might weave unfamiliar new forms—not from paper, and not from pixels, but from insights about human cognition.

Why lectures don't work

We've been discussing books so far, but have you ever had the same type of experience with a *lecture?* It's easy to attend a lecture and feel that you understand, only to discover over that night's problem set that you understood very little. Memory feels partly to blame: you might sense that you knew certain details at one time, but you've forgotten. Yet we can't pin this all on memory. When you pull on certain strings from the lecture, you might discover that *you had never really understood*, though you'd certainly *thought* you understood during the lecture.

You've probably internalized the notion that lectures have this problem, even if the parallel claim for books feels more alien. After all, doesn't this portrait of a teacher sound like a cliché?

> Mr. Johnson talks at his class for an hour each day, believing his students have absorbed every word—then wonders why their test scores are so poor.

Like books, lectures can be *entertaining* or *influential*; like books, lectures do seem to work... sometimes, for some people. But you probably don't believe that lectures are a reliable way to convey knowledge.

Books don't work for the same reason that lectures don't work: neither medium has any explicit theory of how people actually learn things, and as a result, both mediums accidentally (and mostly invisibly) evolved around a theory that's plainly false.

To illustrate what I mean, I'll try to draw on your own learning experiences. You've probably discovered that certain strategies help you absorb new ideas: solving interesting problems, writing chapter summaries, doing creative projects, etc. Whatever strategies you prefer, they're not magic. There's a *reason* they work (when they do): they're leveraging some underlying *truth* about your cognition—about the way you think and learn. In many cases, the truth is not just about your cognition but about human cognition in general.

If we collect enough of these underlying "truths," some shared themes might emerge, suggesting a more coherent theory of how learning happens. We'll call such theories *cognitive models*. Some learning strategies suggest the same model; others suggest conflicting models. Some of these

models are empirically testable; others aren't; still others are already known to be false. By focusing on these models, instead of a herd of one-off strategies, we can seek more general implications. We can ask: if we take a particular cognitive model seriously, what does it suggest will (or won't) help us understand something?

That's an important question because *it's hard to convey knowledge*. Most lecture attendees don't absorb the intended knowledge; most book readers don't absorb the intended knowledge. Failure is the default here. So if you hope to help others understand things, you had better draw on some great ideas about how people learn. It would be nice if this weren't true. It would be nice if one could simply explain an idea clearly to someone, then trust that they've understood it. Unfortunately, as you've likely seen in classrooms and in your own life, complex ideas are rarely understood so automatically.

Lectures, as a medium, have no carefully-considered cognitive model at their foundation. Yet if we were aliens observing typical lectures from afar, we might notice the *implicit* model they appear to share: "the lecturer says words describing an idea; the class hears the words and maybe scribbles in a notebook; then the class understands the idea." In learning sciences, we call this model "transmissionism."

It's the notion that knowledge can be directly transmitted

from teacher to student, like transcribing text from one page onto another. If only! The idea is so thoroughly discredited that "transmissionism" is only used pejoratively, in reference to naive historical teaching practices. Or as an ad-hominem in juicy academic spats.

Of course, good lecturers don't usually *believe* that simply telling their audience about an idea causes them to understand it. It's just that lectures, as a format, are shaped *as if* that were true, so lecturers mostly *behave* as if it were true.

If pressed, many lecturers would offer a more plausible cognitive model: understanding actually comes *after* the lecture, when attendees solve problem sets, write essays, etc. The lecture provides the raw information for those later activities. Great: that's a real model, and parts of it are supported by cognitive science. But if we'd *begun* with this model, would we have chosen live, ninety-minute speeches to convey raw information for a problem set?

Listeners' attention wanders after a few minutes, so wouldn't we want to interleave the problem-solving sessions with the lecture? Live speeches can't be paused or rewound, so aren't they awfully lossy for conveying raw information? People can read much more quickly than a lecturer speaks, so wouldn't text be more efficient? And so on—it's already clear that the traditional lecture format isn't particularly informed by this

model.

The lectures-as-warmup model is a post-hoc rationalization, but it does gesture at a deep theory about cognition: *to understand something, you must actively engage with it*. That notion, taken seriously, would utterly transform classrooms. We'd prioritize activities like interactive discussions and projects; we'd deploy direct instruction only when it's the best way to enable those activities. I'm not idly speculating: for the last few decades, this has been one of the central evolutionary forces in US K–12 policy and practice.

In summary: lectures don't work because the medium lacks a functioning cognitive model. It's (implicitly) built on a faulty idea about how people learn—transmissionism—which we can caricaturize as "lecturer says words describing an idea; students hear words; then they understand." When lectures *do* work, it's generally as part of a broader learning context (e.g. projects, problem sets) with a better cognitive model. But the lectures aren't pulling their weight. If we really wanted to adopt the better model, we'd ditch the lectures, and indeed, that's what's been happening in US K–12 education.

With our intuitions primed through lectures, we'll see that books, as a medium, reflect the same flawed ideas about how people learn.

Why books don't work

Like lectures, books have no carefully-considered cognitive model at their foundation, but the medium does have an *implicit* model. And like lectures, *that model is transmissionism*. Sequences of words in sequences of lines in sequences of pages, the form of a book suggests *people absorb knowledge by reading sentences*. In caricature: "The author describes an idea in words on the page; the reader reads the words; then the reader understands the idea. When the reader reaches the last page, they've finished the book." Of course, most authors don't believe that people learn things this way, but because the medium makes the assumption invisible, it's hard to question.

Like lecturers, many authors would offer a more plausible cognitive model when pressed. Readers can't just *read* the words. They have to really *think* about them. Maybe take some notes. Discuss with others. Write an essay in response. Like a lecture, a book is a warmup for the thinking that happens later. Great: that's a better model! Let's look at how it plays out.

I acknowledged earlier that of course, some people *do* absorb knowledge from books. Indeed, those are the people who really do think about what they're reading. The process is often invisible. These readers' inner monologues have sounds like: "This idea reminds me of...," "This point conflicts

with…," "I don't really understand how…," etc. If they take some notes, they're not simply transcribing the author's words: they're summarizing, synthesizing, analyzing.

Unfortunately, these tactics don't come easily. Readers must learn specific reflective strategies. "What questions should I be asking? How should I summarize what I'm reading?" Readers must run their own feedback loops. "Did I understand that? Should I re-read it? Consult another text?" Readers must understand their own cognition. "What does it feel like to understand something? Where are my blind spots?"

These skills fall into a bucket which learning science calls "metacognition." The experimental evidence suggests that it's challenging to learn these types of skills, and that many adults lack them. [3]Baker, L. (1989). Metacognition, comprehension monitoring, and the adult reader. Educational Psychology Review, 1(1), 3–38. Worse, even if readers know *how* to do all these things, the process is quite taxing. Readers must juggle both the content of the book and also all these meta-questions. People particularly struggle to multitask like this when the content is unfamiliar [4]See e.g. Langer, J. A., & Nicolich, M. (1981). Prior knowledge and its relationship to comprehension. Journal of Reading Behavior, 13(4). and Baker, L., & Brown, A. L. (1984).

Metacognitive skills and reading. Handbook of reading

research, 1(353), V394..

Where is the book in all this? If we believe that successful reading requires engaging in all this complex metacognition, how is that reflected in the medium? What's it doing to help?

Of course, great authors earnestly want readers to think carefully about their words. These authors form sophisticated pictures of their readers' evolving conceptions. They anticipate confusions readers might have, then shape their prose to acknowledge and mitigate those issues. They make constant choices about depth and detail using these models. They suggest what background knowledge might be needed for certain passages and where to go to get it.

By shouldering some of readers' self-monitoring and regulation, these authors' efforts can indeed lighten the metacognitive burden. But metacognition is an inherently dynamic process, evolving continuously as readers' own conceptions evolve. Books are static. Prose can frame or stimulate readers' thoughts, but prose can't behave or respond to those thoughts as they unfold in each reader's head. The reader must plan and steer their own feedback loops.

If lecturers believe that lectures are a warm-up for the understanding developed through problem sets and essays,

then at least the lecturers design those activities and offer feedback on students' work. By comparison, if authors believe that understanding comes only when readers really think about their words, then they're largely leaving readers to design their own "problem sets" and to generate their own feedback. All this effortful "thinking about thinking" competes with actually thinking about the book's ideas. [5] I've oversimplified here a bit. In fact, this kind of meta-processing of material—designing one's own questions and generating one's own feedback—*are* sometimes effective cognitive strategies. But as far as learning science understands it, they're only effective for people who are already proficient with both the object-level concepts and also the relevant metacognitive skills. For others, these activities appear to detract from understanding the material; see e.g. Kalyuga, S. (2009). Knowledge elaboration: A cognitive load perspective. Learning and Instruction, 19(5), 402–410.

If the model is that people understand written ideas by thinking carefully about them, what would books look like if they were built around helping people do that?

What about textbooks?

Wait—isn't that what textbooks do? Can we just slap some exercises and discussion questions on *The Selfish Gene*? It

doesn't exactly sound pleasant to read, but would it work?

Unlike most non-fiction books, textbooks usually *are* built around explicit cognitive models. For instance, they often alternate between explanations which introduce concepts, and exercises which push students to think about those concepts in specific ways. It's great that textbooks didn't choose their cognitive models by accident. That's an important first step. But it's not enough: people still struggle to reliably absorb knowledge from textbooks.

I'll now make the case that textbooks don't effectively implement their own models about how people learn—and that even if they did, textbooks' models neglect important ideas about how people learn.

Let's begin by looking at textbooks in practice. It's striking that academic courses are often structured around textbooks, but lots of people spend the extra time and money to enroll in those courses—rather than just studying the textbooks independently. Indeed, I suspect that textbooks are mostly purchased for course syllabi, not for self-study. Sure: some people take courses because they want a credential. But plenty of students genuinely feel they'll learn more by taking courses than they would by studying those courses' textbooks. Assuming students' feelings aren't completely misplaced, courses must be offering something extra that's important to how people

learn.

We saw earlier how non-fiction books' accidental cognitive model left readers doing all the metacognitive work to plan, execute, and monitor their engagement with the book's ideas. By contrast, textbooks do have explicit cognitive models: they support engagement with their concepts through things like exercises and discussion questions. Yet much of the metacognitive burden still remains with the reader.

Readers must decide which exercises to do and when. Readers must run their own feedback loops: did they clearly understand the ideas involved in the exercise? If not, what should they do next? What should students do if they're completely stuck? Some issues are subtler. For example, textbook exercises are often designed to yield both a solution to that specific problem and also broader insights about the subject. Will readers notice if they solved a problem but missed the insights it was supposed to reveal?

By contrast, courses handle much of this metacognitive burden. Their syllabi offer a scheduled scope and sequence, so students need do less planning of their own. Students generally receive feedback on exercises, both individually and through class-wide discussion. If students are stuck, they can attend office hours to receive finer-grained help. Instructors can discuss the implications of the previous

week's exercises in class. Certainly, courses do none of this perfectly. Plenty of students still absorb nothing from a class. But by shouldering some of the metacognition, courses preserve more of students' attention for the material itself.

At this point, a typical narrative in educational technology would observe how AI-based learning systems [6] For instance, intelligent tutoring systems have for decades specifically aspired to shoulder more task-oriented metacognitive burden. could offer automated feedback and task planning outside of the classroom. There's been intriguing progress here, and these methods can indeed improve on textbooks, but these systems generally fixate on a narrow, task-oriented view of what's happening in classrooms. Academic courses offer more than just metacognitive support for textbooks; their cognitive model is also social and emotional.

For instance, class discussions support social learning: students understand topics more deeply by grappling with their peers' understandings of the same ideas. Courses can provide a personal relationship with a disciplinary expert, a rich conduit for accessing the discipline's culture—much of which may be tacit. For many students, courses offer a helpful accountability structure, playing an important role in supporting their willpower.

Courses also offer emotional salience, which motivates and

amplifies learning: live lectures might be inefficient, but an instructor's palpable fascination can leave a lasting impression. That same pathos saturates the best non-fiction prose, but textbooks generally neglect emotional connection; their prose inspires indifference more often than wonder. Because they typically lack an authorial voice, and because of their obsession with evaluation, computerized tutoring systems have largely produced works with even less regard for pathos.

In this section we've seen that, like lectures, non-fiction books don't work because they lack a functioning cognitive model. Instead, like lectures, they're (accidentally, invisibly) built on a faulty idea about how people learn: transmissionism. When books *do* work, it's generally for readers who deploy skillful metacognition to engage effectively with the book's ideas. This kind of metacognition is unavailable to many readers and taxing for the rest. Books aren't pulling their weight. Textbooks do more to help, but they still foist most of the metacognition onto the reader, and they ignore many important ideas about how people learn.

What to do about it

How might we make books actually work reliably? At this

point, the slope before us might feel awfully steep. Some early footholds might be visible—a few possible improvements to books, or tools one might make to assist readers—but it's not at all clear how to reach the summit. In the face of such a puzzle, it's worth asking: are we climbing the right hill? Why are we climbing this particular hill at all?

I argued earlier that books, as a medium, weren't built around any explicit model of how people learn. It's possible that, in spite of this "original sin," iterative improvements to the form, along with new tools to support readers, can make books much more reliable. But it's also possible that we'll never discover the insights we need while tethered to the patterns of thought implicit in this medium.

Instead, I propose: we don't necessarily have to make books work. *We can make new forms instead*. This doesn't have to mean abandoning narrative prose; it doesn't even necessarily mean abandoning paper—rather, we can free our thinking by abandoning our preconceptions of what a book is. Maybe once we've done all this, we'll have arrived at something which does indeed look much like a book. We'll have found a gentle path around the back of that intimidating slope. Or maybe we'll end up in different terrain altogether.

So let's reframe the question. Rather than "how might we make books actually work reliably," we can ask: How might we design mediums which do the job of a non-fiction book

—but which actually work reliably?

I'm afraid that's a research question—probably for several lifetimes of research—not something I can directly answer in these brief notes. But I believe it's possible, and I'll now try to share why.

To begin, it's important to see that mediums can be designed, not just inherited. What's more: *it is possible to design new mediums which embody specific ideas*. Inventors have long drawn on this unintuitive insight [7]See e.g. Douglas Engelbart's 1962 "Augmenting Human Intellect" for a classic primary source or Michael Nielsen's 2016 "Thought as a Technology" for a synthesis of much work in this space., but I'll briefly review it in case it's unfamiliar. Mathematical proofs are a medium; the step-by-step structure embodies powerful ideas about formal logic. Snapchat Stories are a medium; the ephemerality embodies powerful ideas about emotion and identity. The World Wide Web is a medium (or perhaps many mediums); the pervasive hyperlinks embody powerful ideas about the associative nature of knowledge.

Perhaps most remarkably, the powerful ideas are often invisible: it's not like we generally think about cognition when we sprinkle a blog post with links. But the people who created the Web *were* thinking about cognition. They designed its building blocks so that the natural way of reading and writing in this medium would reflect the

powerful ideas they had in mind. Shaped intentionally or not, each medium's fundamental materials and constraints give it a "grain" which make it bend naturally in some directions and not in others.

This "grain" is what drives me when I gripe that books lack a functioning cognitive model. It's not just that it's possible to create a medium *informed* by certain ideas in cognitive science. Rather, it's possible to weave a medium *made out of* those ideas, in which a reader's thoughts and actions are inexorably—perhaps even invisibly—shaped by those ideas. Mathematical proofs, as a medium, don't just *consider* ideas about logic; we don't *attach* ideas about logic to proofs. *The form is made out of ideas about logic.*

How might we design a medium so that its "grain" bends in line with how people think and learn? So that by simply engaging with an author's work in the medium—engaging in the obvious fashion; engaging in this medium's equivalent of books' "read all the words on the first page, then repeat with the next, and so on"—one would automatically do what's necessary to understand? So that, in some deep way, the default actions and patterns of thought when engaging with this medium *are the same thing as* "what's necessary to understand"?

That's a tall order. Even on a theoretical level, it's not clear what's necessary for understanding. Indeed, that framing's

too narrow: there are many paths to understanding a topic. But cognitive scientists and educators have mapped some parts of this space, and they've distilled some powerful ideas we can use as a starting point.

For example, people struggle to absorb new material when their working memory is already overloaded. More concretely: if you've just been introduced to a zoo of new terms, you probably won't absorb much from a sentence which uses many of those terms at once. So maybe part of "what's necessary to understand" something is that most of its prerequisites must be not just familiar but fluent, encoded in long-term memory.

To help people encode more into long-term memory, we can draw on another powerful idea from cognitive science: spaced repetition. By re-testing yourself on material you've learned over expanding intervals, you can cheaply and reliably commit huge volumes of information to long-term memory [8]For a review of this effect's practical implications, see Michael Nielsen's "Augmenting Long-term Memory". For more on the empirical evidence, this review by David Balota and colleagues is a good starting point.. Of course, memory is only a small slice of "understanding," but to illustrate how one might begin to address understanding as a whole, let's explore how we might weave a medium out of these two

ideas about memory.

My collaborator Michael Nielsen and I made an initial attempt with *Quantum Country*, a "book" on quantum computation. But reading this "book" doesn't look like reading any other book. The explanatory text is tightly woven with brief interactive review sessions, meant to exploit the ideas we just introduced. Reading *Quantum Country* means reading a few minutes of text, then quickly testing your memory about everything you've just read, then reading for a few more minutes, or perhaps scrolling back to reread certain details, and so on. Reading *Quantum Country* also means repeating those quick memory tests in expanding intervals over the following days, weeks, and months. If you read the first chapter, then engage with the memory tests in your inbox over the following days, we expect your working memory will be substantially less taxed when reading the second chapter. What's more, the interleaved review sessions lighten the metacognitive burden normally foisted onto the reader: they help readers see where they're absorbing the material and where they're not.

Quantum Country is just one piece of the memory puzzle, which itself is part a larger tapestry. How might we design mediums in which "readers" naturally form rich associations between the ideas being presented? How might we design mediums which "readers" naturally engage creatively with

the material? How might we design mediums in which "readers" naturally contend with competing interpretations? If we pile together enough of these questions we're left with: how might we design mediums in which "reading" is the same as "understanding"? A more detailed treatment of such a research program is beyond the scope of these brief notes, but I believe that the answers to questions like these can transform the pace of human knowledge, echoing the transformation which books themselves sparked so long ago.

Want to hear when I publish new work?

Your email, please

SUBSCRIBE

Acknowledgements

here, my own views are deeply influenced by Michael's perspectives.

As always, my acknowledgment of these kind contributions should not be construed as endorsements on the part of these individuals.

Citing this work

In academic work, please cite this as:

License

History

This note was first published on May 11, 2019.