

# How can we develop transformative tools for thought?

Andy Matuschak and Michael Nielsen

Part of the origin myth of modern computing is the story of a golden age in the 1960s and 1970s. In this story, visionary pioneers pursued a dream in which computers enabled powerful tools for thought, that is, tools to augment human intelligence. One of those pioneers, Alan Kay, summed up the optimism of this dream when he wrote of the potential of the personal computer: “the very use of it would actually change the thought patterns of an entire civilization”.

It’s an inspiring dream, which helped lead to modern interactive graphics, windowing interfaces, word processors, and much else. But retrospectively it’s difficult not to be disappointed, to feel that computers have not yet been nearly as transformative as far older tools for thought, such as language and writing. Today, it’s common in technology circles to pay lip service to the pioneering dreams of the past. But nostalgia aside there is little determined effort to pursue the vision of transformative new tools for thought.

We believe now is a good time to work hard on this vision again. In this essay we sketch out a set of ideas we believe can be used to help develop transformative new tools for thought. In the first part of the essay we describe an experimental prototype system that we’ve built, a kind of *mnemonic medium* intended to augment human memory. This is a snapshot of an ongoing project, detailing both encouraging progress as well as many challenges and opportunities. In the second part of the essay, we broaden the focus. We sketch several other prototype systems. And we address the question: why is it that the technology industry has made comparatively little effort developing this vision of transformative tools for thought?

In the opening we mentioned some visionaries of the past. To those could be added many others – Ivan Sutherland, Seymour Papert, Vannevar Bush, and more. Online there is much well-deserved veneration for these people. But such veneration can veer into an unhealthy reverence for the good old days, a belief that giants once

E.g., Douglas Engelbart,  
[Augmenting Human Intellect: A Conceptual Framework](#) (1962).

[Alan Kay, User Interface: A Personal View](#) (1989).

roamed the earth, and today’s work is lesser. Yes, those pioneers did amazing things, and arguably had ways of working that modern technologists, in both industry and academia, are poorly equipped to carry on. But they also made mistakes, and were ignorant of powerful ideas that are available today. And so a theme through both parts of the essay is to identify powerful ideas that weren’t formerly known or weren’t acted upon. Out of this understanding arises a conviction that a remarkable set of opportunities is open today.

A word on nomenclature: the term “tools for thought” rolls off neither the tongue nor the keyboard. What’s more, the term “tool” implies a certain narrowness. Alan Kay has argued that a more powerful aim is to develop a new *medium for thought*. A medium such as, say,

Adobe *Illustrator* is essentially different from any of the individual tools *Illustrator* contains. Such a medium creates a powerful immersive context, a context in which the user can have new kinds of thought, thoughts that were formerly impossible for them. Speaking loosely, the range of expressive thoughts possible in such a medium is an emergent property of the elementary objects and actions in that medium. If those are well chosen, the medium expands the possible range of human thought.

With that said, the term “tools for thought” has been widely used since Iverson’s 1950s and 1960s work introducing the term. And so we shall use “tools for thought” as our catch all phrase, while giving ourselves license to explore a broader range, and also occasionally preferring the term “medium” when it is apt.

An account may be found in Iverson’s Turing Award lecture, [Notation as a Tool of Thought](#) (1979). Incidentally, even Iverson is really describing a medium for thought, the APL programming language, not a narrow tool.

Let’s come back to that phrase from the opening, about changing “the thought patterns of an entire civilization”. It sounds ludicrous, a kind of tech soothsaying. Except, of course, such changes have happened multiple times during human history: the development of language, of writing, and our other most powerful tools for thought. And, for

better and worse, computers really have affected the thought patterns of our civilization over the past 60 years, and those changes seem like just the beginning. This essay is a small contribution to understanding how such changes happen, and what is still possible.

The musician and comedian Martin Mull has observed that “writing about music is like dancing about architecture”. In a similar way, there’s an inherent inadequacy in writing about tools for thought. To the extent that such a tool succeeds, it expands your thinking beyond what can be achieved using existing tools, including writing. The more transformative the tool, the larger the gap that is opened. Conversely, the larger the gap, the more difficult the new tool is to evoke in writing. But what writing can do, and the reason we wrote this essay, is act as a bootstrap. It’s a way of identifying points of leverage that may help develop new tools for thought. So let’s get on with it.

## Part I: Memory systems

### Introducing the mnemonic medium

Few subjects are more widely regarded as difficult than quantum computing and quantum mechanics. Indeed, popular media accounts often regale (and intimidate) readers with quotes from famous physicists in the vein of: “anyone who thinks they’ve understood quantum mechanics has not understood quantum mechanics”.

What makes these subjects difficult? In fact, individually many of the underlying ideas are not too complicated for people with a technical background. But the ideas come in an overwhelming number, a tsunami of unfamiliar concepts and notation. People must learn in rapid succession of qubits, the bra-ket notation, Hadamard gates, controlled-not gates, and many, many other abstract, unfamiliar notions. They’re imbibing an entire new language. Even if they can follow at first, understanding later ideas requires fluency with all the earlier ideas. It’s overwhelming and eventually disheartening.

As an experiment, we have developed a website, *Quantum Country*, which explores a new approach to explaining quantum computing and quantum mechanics. Ostensibly, *Quantum Country* appears to be a conventional essay introduction to these subjects. There is text, explanations, and equations, much as in any other technical essay.

Here’s an excerpt:

In this example, the state  $0.6|0\rangle + 0.8|1\rangle$  is just 0.6 times the  $|0\rangle$  vector, plus 0.8 times the  $|1\rangle$  vector. In the usual vector notation that means the state is:

$$0.6|0\rangle + 0.8|1\rangle = 0.6 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0.8 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}.$$

I’ve been talking about quantum states as two-dimensional vectors. What I didn’t yet mention is that in general they’re *complex vectors*, that is, they can have complex numbers as entries. Of course, the example just shown has real entries, as do the computational basis states. But for a general quantum state the entries can be complex numbers. So, for instance, another quantum state is the vector

$$\frac{1+i}{2}|0\rangle + \frac{i}{\sqrt{2}}|1\rangle,$$

But it’s not a conventional essay. Rather, *Quantum Country* is a prototype for a new type of *mnemonic medium*. Aspirationally, the mnemonic medium makes it almost effortless for users to remember what they read. That may sound like an impossible aspiration. What makes it plausible is that cognitive scientists know a considerable amount about how human beings store long-term memories. Indeed, what they know can almost be distilled to an actionable recipe: follow these steps, and you can remember whatever you choose.

Unfortunately, those steps are poorly supported by existing media. Is it possible to design a new medium which much more actively supports memorization? That is, the medium would build in (and, ideally, make almost effortless) the key steps involved in memory. If we could do this, then instead of memory being a haphazard event, subject to chance, the mnemonic medium would make memory into a choice. Of course, on its own this wouldn’t make it trivial to learn subjects such as quantum mechanics and quantum computing – learning those subjects is about much more than memory. But it would help in addressing one core difficulty: the overwhelming number of new concepts and notation.

For more on this argument, see Andy Matuschak, [Why books don’t work](#) (2019).

In fact, there are many ways of redesigning the essay medium to do that. Before showing you our prototype, please pause for a moment and consider the following questions: how could you build a medium to better support a person’s memory of what they read? What interactions could easily and enjoyably help people consolidate memories? And, more broadly: is it possible to 2x what people remember? 10x? And would that make any long-term difference to their effectiveness?

Composing multiple quantum gates corresponds to the operation of matrix \_\_\_\_\_ with the corresponding matrices.

Click anywhere to reveal Click anywhere to reveal



Didn't remember Remembered

Composing multiple quantum gates corresponds to the operation of matrix \_\_\_\_\_ with the corresponding matrices.

multiplication

in-text 5 days 2 weeks 1 month 2 months long-term

Didn't remember Remembered

Composing multiple quantum gates corresponds to the operation of matrix \_\_\_\_\_ with the corresponding matrices.

multiplication

in-text 5 days 2 weeks 1 month 2 months long-term

Didn't remember Remembered

What is the product of the  $X$  matrix with itself,  $XX$ ?

Click anywhere to reveal Click anywhere to reveal



Didn't remember Remembered

[Click here](#) to view a video of this interaction.

We use the term *cards* for these interface elements pairing questions and answers.

Of course, for long-term memory it's not enough for users to be tested just once on their recall. Instead, a few days after first reading the essay, the user receives an email asking them to sign into a review session. In that review session they're tested again, in a manner similar to what was shown above. Then, through repeated review sessions in the days and weeks ahead, people consolidate the answers to those questions into their long-term memory.

So far, this looks like no more than an essay which integrates old-fashioned flashcards. But notice the intervals indicated at the bottom of the cards:

in-text 5 days 2 weeks 1 month 2 months long-term

Let's sketch the user experience of *Quantum Country*. At the time of this writing the site contains three mnemonic essays (i.e., particular instances of the mnemonic medium). We'll focus on the introductory essay, "Quantum Computing for the Very Curious". Embedded within the text of the essay are 112 questions about that text. Users are asked to create an account, and quizzed as they read on whether they remember the answers to those questions. The figure above shows the interaction as a user answers a question.

Note that this interaction occurs within the text of the essay itself. Here's a zoomed-out view, so you can see how such questions are surrounded by essay text both above and below:

wire. And so we've arrived at the same conclusion in a different way, using matrix multiplication. Doing such matrix multiplications is a pretty common way of analyzing quantum circuits.

Composing multiple quantum gates corresponds to the operation of matrix \_\_\_\_\_ with the corresponding matrices.

Click anywhere to reveal Click anywhere to reveal

Didn't remember Remembered

The highlighted time interval is the duration until the user is tested again on the question. Questions start out with the time interval "in-text", meaning the user is being tested as they read the essay. That rises to five days, if the user remembers the answer to the question. The interval then continues to rise upon each successful review, from five days to two weeks, then a month, and so on. After just five successful reviews the interval is at four months. If the user doesn't remember at any point, the time interval drops down one level, e.g., from two weeks to five days.

This takes advantage of a fundamental fact about human memory: as we are repeatedly tested on a question, our memory of the answer gets stronger, and we are likely to retain it for longer. This exponential rise perhaps seems innocuous, but it's transformative. It means that a relatively small number of reviews will enable a user to remember for years. With

The literature on this effect is vast. A useful entrée is: Gwern Branwen, [Spaced Repetition for Efficient Learning](#).

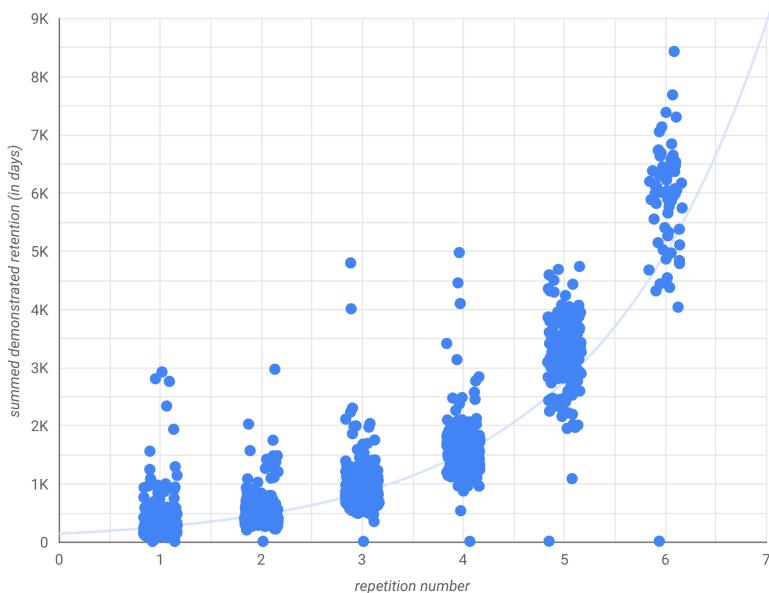
## The Hadamard gate

We've seen our first quantum gate, the NOT or  $X$  gate. Of course, the  $X$  didn't appear to do all that much beyond what is possible with a classical

the time taken to review a typical question being just a few seconds, that means a user can achieve long-term recall with no more than a few minutes' work. By contrast, with conventional flashcards it takes hours of review to achieve the same durability. Exponential scheduling is far more efficient.

## The early impact of the prototype mnemonic medium

Although it's early days for *Quantum Country* we can begin to see some of the impact of the mnemonic medium. Plotted below is the demonstrated retention of answers for each user, versus the number of times each question in the mnemonic essay has been reviewed:



The graph takes a little unpacking to explain. By a card's "demonstrated retention" we mean the maximum time between a successful review of that card, and the prior review of that card. A little more concretely, consider repetition number 6, say (on the horizontal axis). At the point, a user has reviewed all 112 questions in the essay 6 times. And the vertical axis shows the total demonstrated retention, summed across all cards, with each blue dot representing a single user who has reached repetition 6.

So, for instance, after 6 repetitions, we see from the graph that most users are up around 6,000 days of demonstrated retention. That means an average of about  $6,000 / 112 \sim 54$  days per question in the essay. Intuitively, that seems pretty good – if you're anything like us, a couple of months after reading something you have only a hazy memory. By contrast, these users have, at low time cost to themselves (of which more below), achieved nearly two months of demonstrated retention across 112 detailed questions.

Furthermore, you can see the exponential rise in retention with the number of times cards have been reviewed. After the first review, users typically have an average of just over 2 days of demonstrated retention, per card. But by the sixth review that rises to an average of 54 days of demonstrated retention. That typically takes about 95 minutes of total review time to achieve. Given that the essay takes about 4 or so hours to read, this suggests that a less than 50% overhead in time commitment can provide many months or years of retention for almost all the important details in the essay.

Particularly careful readers may wonder how this is possible, given that we stated earlier that the first review interval is 5 days. The explanation is that we recently modified the review schedule so the first review is after 5 days. For most of *Quantum Country*'s history the review schedule was more conservative, and this is the reason for the difference.

This is the big, counterintuitive advantage of spaced repetition: you get exponential returns for increased effort. On average, every extra minute of effort spent in review provides more and more benefit. This is in sharp contrast with most experiences in life, where we run into diminishing returns. For instance, ordinarily if you increase the amount of time you spend reading by 50%, you expect to get no more than 50% extra out of it, and possibly much less. But with the mnemonic medium when you increase the amount of time you spend reading by 50%, you may get 10x as much out of it. Of course, we don't quite mean those numbers literally. But it does convey the key idea of getting a strongly non-linear return. It's a change in the quality of the medium.

This delayed benefit makes the mnemonic medium unusual in multiple ways. Another is this: most online media use short-term engagement models, using variations on operant conditioning to drive user behavior. This is done by Twitter, Facebook, Instagram, and many other popular media forms. The mnemonic medium is much more like meditation – in some ways, the anti-product, since it violates so much conventional Silicon Valley wisdom – in that the benefits are delayed, and hard to have any immediate sense of. Indeed, with the mnemonic medium, the greater the delay, the more the benefit.

These are preliminary results, and need more investigation. One naturally wonders what would happen if we'd been much more aggressive with the review schedule, setting the initial interval between reviews to (say) 2 months? If users reliably retained information up to that point, then the graph would start very high, and we wouldn't see the exponential. We need to investigate these

and many similar questions to better understand what's going on with user's memories.

Early feedback from users makes us cautiously optimistic that they're finding the mnemonic medium useful. In May 2019, one of us posted to Twitter a short thread explaining the technical details of how quantum teleportation works. One user of *Quantum Country* replied to the thread with:

I've only done your first quantum country course (so far) but I find it remarkable that I can view the proof and follow it, knowing what everything means. It's almost like Neo in The Matrix telling Morpheus, 'I know quantum computing'.

A user with significantly more prior experience of quantum computing wrote:

I have a PhD in quantum information/computing and I knew everything in the essay before reading it, but the additional understanding I got from doing the given spaced repetition flashcards significantly improved my understanding of the material. Everyone who is reading this essay, should sign up and give spaced repetition a try.

Another user, new to quantum computing, told us that *Quantum Country* "is by far the best way that I could imagine being introduced to this material". When we asked how he'd used what he'd learned, he explained that when a visitor to his company gave a technical seminar about quantum computing, he expected to get lost after about 10 minutes. Instead:

Wow, I actually followed that for 40 or 45 minutes because the matrices looked familiar... [the medium means] you run into concepts over and over again... It affords interactions at a more effective level of abstraction.

Site analytics show a constant flow of people steadily working through the review sessions in the manner we intended. Six months after release of the prototype, 195 users had demonstrated one full month of retention on at least 80% of cards in the essay, demonstrating an extraordinary level of commitment to the process. We don't yet have a good model of exactly what those people are learning, but it seems plausible they are taking away considerably more than from a conventional essay, or perhaps even from a conventional class.

Of course, this kind of feedback and these kinds of results should be taken with a grain of salt. The mnemonic medium is in its early days, has many deficiencies, and needs improvement in many ways (of which more soon). It is, however, encouraging to hear that some users already find the medium exceptionally helpful, and suggests developing and testing the medium further. At a minimum, it seems likely the mnemonic medium is genuinely helping people remember. And furthermore it has the exponentially increasing efficiency described above: the more people study, the more benefit they get per minute studied.

In another informal experiment, we tried to figure out how much it affected user's memories when they weren't asked to review cards. To do this, we introduced a deliberate short (two-week) delay on reviews for a small subset of 8 cards. That is, some users would review those 8 cards upon an initial read, and then would be prevented from reviewing them again for at least two weeks. Other users would continue to study as normal on the 8 cards. By comparing the two groups we could estimate the effect that reviewing the cards had on user's memories.

What happened? Well, for those users whose reviews were delayed, accuracy dropped from 91% (upon the initial read) to 87% (after two weeks). This may seem a small drop, but keep in mind that users continued to review other cards, which almost certainly boosted their final performance, since those other cards had some overlap in content with the delayed cards. It's difficult to avoid this kind of overlap without delaying reviews on all cards, a more drastic change in user experience than we wanted to impose. For users who were asked to review the cards as normal, accuracy improved from 89% to 96%. The short summary is: when users didn't review the cards, accuracy dropped by 4%; when they did review the cards, accuracy increased by 7%.

Another way of looking at the data from this informal experiment is to ask which users saw improved or unchanged performance, and which saw their performance get worse. In fact, every single user (100%) who reviewed cards on the regular schedule saw

In more detail: there were 16 users in the group that did the reviews, per usual, and 25 users in the group where reviews were delayed. The 95% confidence intervals were:  $91 \pm 4\%$ ,  $87 \pm 5\%$ ,  $89 \pm 5\%$ ,  $96 \pm 3\%$ , assuming each variable is binomial, independent and identically distributed. This latter assumption is approximate, since we'd expect some user- and question-dependent effects. Note also that this was done in an earlier version of *Quantum Country*, where the review schedule had intervals of one day, three days, one week, and two weeks.

We've made no attempt at all to scale this out. It's interesting to ponder doing so.

their performance either stay the same or improve. By contrast, 40% of the users whose reviews were delayed saw their performance get worse, while 60% saw it stay the same or improve.

These are small-but-promising results. Of course, our experiment was only done over two weeks, and we'd expect larger effects in experiments done over longer periods. And, as already mentioned, the effect is likely diminished by overlaps between the cards. Nonetheless, this informal experiment again suggests the mnemonic medium is helping people's memory, and suggests more comprehensive studies.

Despite these suggestive preliminary results, it's still tempting to be dismissive. Isn't this "just" an essay with flashcards embedded? At some level, of course, that's correct. In the same way, wikis are just editable web pages; Twitter is just a way of sharing very short form writing; and Facebook is just a way of sharing writing and pictures with friends. Indeed, writing itself is just a clever way of ordering a small number of symbols on a page. While a medium may be simple, that doesn't mean it's not profound. As we shall see, the mnemonic medium has many surprising properties. It turns out that flashcards are dramatically under-appreciated, and it's possible to go much, much further in developing the mnemonic medium than is *a priori* obvious.

Before we delve deeper into the mnemonic medium, let's mention one challenge in the discussion: the inherent difficulty in achieving a good balance between conveying enthusiasm and the kind of arm's-length skepticism appropriate for evaluation. On the one hand, we would not have built the mnemonic medium if we weren't excited about the underlying ideas, and wanted to develop those enthusiasms. To explain the mnemonic medium well, we need to bring you, the reader, inside that thinking. But having done that, we also need to step back and think more skeptically about questions such as: is this medium really working? What effect is it actually having on people? Can it be made 10x better? 100x better? Or, contrariwise, are there blockers that make this an irredeemably bad or at best mediocre idea? How important a role does memory play in cognition, anyway? So far, we've focused on the enthusiastic case for the medium, why one might consider this design at all. But later in this essay we'll gradually step back and reflect in a more skeptical frame.

## Expanding the scope of memory systems: what types of understanding can they be used for?

*Quantum Country* is an example of a *memory system*. That is, it's a system designed to help users easily consolidate what they've learned into long-term memory. It's part of a long history of memory systems, going back to ancient times, when the orator Cicero and the rhetorician Quintilian described mnemonic techniques that could be used to memorize long texts.

In modern times, many memory systems have been developed. Among the better known are Anki, SuperMemo, Quizlet, Duolingo, and Memrise. Like *Quantum Country*, each of these systems uses increasing time intervals between reviews of particular questions. Such systems are sometimes known as *spaced-repetition memory systems* (or *SRM systems*). They're usually justified in a manner similar to our explanation for *Quantum Country*: some notion of each review gradually increasing the consolidation strength for a memory.

SRM systems are most widely used in language learning. Duolingo, for instance, claims [25 million monthly active users](#). Reports are mixed on success. Some serious users are [enthusiastic about their success](#) with Duolingo. But [others find it of limited utility](#). The company, of course, [touts research](#) showing that it's incredibly successful. It seems likely to us that Duolingo and similar systems are useful for many users as part of (but only part of) a serious language learning program.

What about memory systems for uses beyond language? Quizlet is popular, with [50 million monthly active users](#). It's widely used in classrooms, especially for simple declarative knowledge – lists of American Presidents, capitals of countries, and so on. Anki and SuperMemo seem to most often be used for similar simple declarative knowledge, but have much smaller active user bases than Quizlet.

One of the ideas motivating *Quantum Country* is that memory systems aren't just useful for simple declarative knowledge, such as vocabulary words and lists of capitals. In fact, memory systems can be extraordinarily helpful for mastering abstract, conceptual knowledge, the kind of knowledge required to learn subjects such as quantum mechanics and quantum computing. This is achieved in part through many detailed strategies for constructing cards capable of encoding this kind of understanding. But, more importantly, it's possible because of the way the mnemonic medium embeds spaced repetition inside a

Strictly speaking, Quizlet's basic product doesn't use spaced repetition. There is, however, a paid version using spaced repetition, and it's otherwise quite similar to many of these systems.

They are, however, widely used within some interesting niche audiences. For instance, there is a [thriving population](#) of medical students using Anki.

narrative. That narrative embedding makes it possible for context and understanding to build in ways difficult in other memory systems.

Other people have also developed ways of using memory systems for abstract, conceptual knowledge. Perhaps most prominently, the creator of the SuperMemo system, Piotr Wozniak, has [written extensively](#) about the many ingenious ways he uses memory systems. And several other expert users of memory systems have also developed similar strategies. However, employing those strategies requires considerable skill. In practice, that skill barrier has meant these strategies are used by no more than a tiny handful of people.

By contrast, in *Quantum Country* an expert writes the cards, an expert who is skilled not only in the subject matter of the essay, but also in strategies which can be used to encode abstract, conceptual knowledge. And so *Quantum Country* provides a much more scalable approach to using memory systems to do abstract, conceptual learning. In some sense, *Quantum Country* aims to expand the range of subjects users can comprehend at all. In that, it has very different aspirations to all prior memory systems.

More generally, we believe memory systems are a far richer space than has previously been realized. Existing memory systems barely scratch the surface of what is possible. We've taken to thinking of *Quantum Country* as a *memory laboratory*. That is, it's a system which can be used both to better understand how memory works, and also to develop new kinds of memory system. We'd like to answer questions such as:

- What are new ways memory systems can be applied, beyond the simple, declarative knowledge of past systems?
- How deep can the understanding developed through a memory system be? What patterns will help users deepen their understanding as much as possible?
- How far can we raise the human capacity for memory? And with how much ease? What are the benefits and drawbacks?
- Might it be that one day most human beings will have a regular *memory practice*, as part of their everyday lives? Can we make it so memory becomes a choice; is it possible to in some sense solve the problem of memory?

Over the next few sections we sketch out some of our thinking about how memory systems may be developed. We'll see that memory systems are a small part of a much bigger picture. Not only is seriously developing memory systems likely to lead to one or more transformative tools for thought, we also believe it will teach us much about the general problem of developing such tools.

## Improving the mnemonic medium: making better cards

In writing mnemonic essays, it's tempting to treat the content of the cards rather casually. After all, a card is just a question and an answer, each containing a little text, perhaps a figure. Surely they ought to be easy to write?

While thinking in this way is tempting, it's a mistake. In fact, cards are fundamental building blocks of the mnemonic medium, and card-writing is better thought of as an open-ended skill. Do it poorly, and the mnemonic medium works poorly. Do it superbly well, and the mnemonic medium can work very well indeed. By developing the card-writing skill it's possible to expand the possibilities of the medium.

A helpful comparison is to the sentence in written prose. For the beginning writer it's tempting to treat sentences casually. But in the hands of a great writer – say, a Nabokov – sentences can be developed into a virtuoso artform. What would it take to achieve virtuoso skill in writing the cards of the mnemonic medium?

It's not obvious *a priori* that writing cards is such a rich activity. One of us wrote 17,000- and 6,000-word essays whose subject was in large part understanding how to write good cards. He didn't realize that was going to be the subject when he began writing; it only became clear in retrospect how rich card writing is. It turns out that answering the question "how to write good cards?" requires thinking hard about your theory of knowledge and how to represent it, and your theory of learning. The better those theories, the better your cards will be. Small wonder it's a rich, open-ended problem!

Michael Nielsen, [Augmenting Long-term Memory](#) (2018), and Michael Nielsen, [Using spaced repetition systems to see through a piece of mathematics](#) (2019).

All that said, let's make a few concrete observations about good card-writing. While the specific examples that follow are relatively banal, they should give you some feeling for the profound issues that arise in improving the mnemonic medium. We'll begin with three principles we used when writing the cards in *Quantum Country*. Note that these are just three of many more principles – a more detailed discussion of good principles of card construction may be found in [Augmenting Long-term Memory](#).

- **Most questions and answers should be atomic:** Early in his own personal memory practice, one of us was learning the Unix command to create links in the filesystem. He entered the following question into his memory system: “How to create a soft link from linkname to filename”. Together with the corresponding answer “ln -s filename linkname”. This looks like a good question, but he routinely forgot the answer. To address this, he refactored the card into two more atomic cards. One card: “What’s the basic command and option to create a soft link?” (A: “ln -s”). Second card: “When creating a soft link, in what order do linkname and filename go?” (A: “filename linkname”). Breaking the card into more atomic pieces turned a question he routinely got wrong into two questions he routinely got right. It seemed that the more atomic questions brought more sharply into focus what he was forgetting, and so provided a better tool for improving memory. And what of the original card? Initially, he deleted it. But he eventually added the card back, with the same question and answer, since it served to integrate the understanding in the more atomic cards.

- **Make sure the early questions in a mnemonic essay are trivial:** it helps many users realize they aren’t paying enough

attention as they read: This was a discovery made when we released the first *Quantum Country* essay.

Anticipating that users would be struggling with a new interface, we deliberately made the first few questions in the essay utterly trivial – sort of a quantum equivalent to “ $2+2 = ?$ ” – so they could focus on the interface. To our surprise, users performed poorly on these questions, worse than they did on the (much harder) later questions. Our current hypothesis to explain this is that when users failed to answer the first few questions correctly it served as a wakeup call. The questions were so transparently simple that they realized they hadn’t really been paying attention as they read, and so were subsequently more careful.

- **Avoid orphan cards:** These are cards which don’t connect closely to anything else. Suppose, for the sake of illustration, that you’re trying to learn about African geography, and have a question: “What’s the territory in Africa that Morocco disputes?” (A: “The Western Sahara”) If you don’t know anything about the Western

Sahara or Morocco or why there’s a dispute, that question will be an orphan, disconnected from everything else. Ideally, you’ll have a densely interconnected web of questions and answers, everything interwoven in striking ways.

Ultimately, we’d like to distill out a set of useful practical principles and idioms to help write good cards and, more generally, good mnemonic essays. Aspirationally, such a set of principles and idioms would work much like *The Elements of Style* (or some similar book of prose advice), and would help other people learn to write high-quality mnemonic essays.

When we first described *Quantum Country* above we explained it using a simple model of spaced repetition: increased consolidation strength for memories leading to increased time intervals between reviews. This is a helpful simple model, but risks creating the misleading impression that it’s all that’s going on in the system. In fact, for the mnemonic medium to work effectively, spaced repetition must be deployed in concert with many other ideas. The three ideas we just described – atomicity of questions and answers, making early questions trivial, avoiding orphan cards – are just three of dozens of important ideas used in the mnemonic medium. We won’t enumerate all those other ideas here – that’s not the purpose of this essay. But we want to emphasize this point, since it’s common for people to have the simplistic model “good memory system = spaced repetition”. That’s false, and an actively unhelpful way of thinking.

Indeed, thinking in this way is one reason spaced-repetition memory systems often fail for individuals. We often meet people who say “Oh, I thought spaced repetition sounded great, and I tried Anki [etc], but it doesn’t work for me”. Dig down a little, and it turns out the person is using their memory system in a way guaranteed to fail. They’ll be writing terrible questions, or using it to learn a subject they don’t care about, or making some other error. They’re a little like a person who thinks “learning the guitar sounds great”, picks it up for half an hour, and then puts it down, saying that they sound terrible and therefore it’s a bad instrument. Of course, what’s really going on is that the guitar and memory systems are both skills that take time to develop. But, with that said, we want to build as much support as possible into the medium. Ideally, even novices would benefit tremendously from the mnemonic medium. That means building in many ideas that go beyond the simplistic model of spaced repetition.

One of us has previously

Michael Nielsen, [Augmenting Long-Term Memory](#) (2018).

asserted that in spaced-repetition memory systems, users need to make their own cards. The reasoning is informal: users often report dissatisfaction and poor results when working with cards made by others. The reason seems to be that making the cards is itself an important act of understanding, and helps with committing material to memory. When users work with cards made by others, they lose those benefits.

*Quantum Country* violates this principle, since users are not making the cards. This violation was a major concern when we began working on *Quantum Country*. However, preliminary user feedback suggests it has worked out adequately. A possible explanation is that, as noted above, making good cards is a difficult skill to master, and so what users lose by not making their own cards is made up by using what are likely to be much higher-quality cards than they could have made on their own. In future, it's worth digging deeper into this issue, both to understand it beyond informal models, and to explore ways of getting the benefits of active card making.

Above we discussed three principles of good question-and-answer construction. Of course, it's also possible to make more structural modifications to the nature of the cards themselves. Here's three questions suggesting experiments in this vein:

- **How can we ensure users don't just learn surface features of questions?** One question in *Quantum Country* asks: "Who has made progress on using quantum computers to simulate quantum field theory?" with the answer: "John Preskill and his collaborators". This is the only "Who...?" question in the entire essay, and many users quickly learn to recognize it from just the "Who...?" pattern, and parrot the answer without engaging deeply with the question. This is a common failure mode in memory systems, and it's deadly to understanding. One response, which we plan to trial soon, is to present the question in multiple different-but-equivalent forms. So the user first sees the question as "Who has made progress [etc]?", but then the second time the question is presented as a fill-in-the-blanks: "\_\_\_ and his collaborators have made progress on using quantum computers to simulate quantum field theory." And so on, multiple different forms of the question, designed so the user must always engage deeply with the meaning of the question, not its superficial appearance. Ultimately, we'd like to develop a library of techniques for identifying when this learning-the-surface-feature pattern is occurring, and for remedying it.

- **How to best help users when they forget the answer to a question?** Suppose a user can't remember the answer to the question: "Who was the second President of the United States?" Perhaps they think it's Thomas Jefferson, and are surprised to learn it's John Adams. In a typical spaced-repetition memory system this would be dealt with by decreasing the time interval until the question is reviewed again. But it may be more effective to follow up with questions designed to help the user understand some of the surrounding context. E.g.: "Who was George Washington's Vice President?" (A: "John Adams"). Indeed, there could be a whole series of followup questions, all designed to help better encode the answer to the initial question in memory.

- **How to encode stories in the mnemonic medium?** People often find certain ideas most compelling in story form. Here's a short, fun example: did you know that Steve Jobs actively opposed the development of the App Store in the early days of the iPhone? It was instead championed by another executive at Apple, Scott Forstall. Such a story carries a force not carried by declarative facts alone. It's one thing to know in the abstract that even the visionaries behind new technologies often fail to see many of their uses. It's quite another to hear of Steve Jobs arguing with Scott Forstall against what is today a major use of a technology Jobs is credited with inventing. Can the mnemonic medium be used to help people internalize such stories? To do so would likely violate the principle of atomicity, since good stories are rarely atomic (though this particular example comes close). Nonetheless, the benefits of such stories seem well worth violating atomicity, if they can be encoded in the cards effectively.

It's easy to generate dozens more questions and ideas in a similar vein. The mnemonic medium is not a fixed form, but rather a platform for experimentation and continued improvement.

One useful metaphor for thinking about how to improve the mnemonic medium is to think of each mnemonic essay as a conventional essay accompanied by a kind of "reflected essay" – the knowledge encoded by all the cards. A user can, with ease, choose to remember as much of that reflected essay as they wish. Of course, the reflection is imperfect. But by developing good card-making strategies we can make the reflected essay a nearly faithful reflection of all the important ideas, the ideas a reader would ideally like to retain.

We said above that it's a mistake to use the simplistic model "good memory system = spaced repetition". In fact, while spaced repetition is a helpful way to introduce *Quantum Country*, we certainly shouldn't pigeonhole the mnemonic medium inside the paradigm of existing SRM systems. Instead, it's better to go back to first principles, and to ask questions like: what would make *Quantum Country* a good memory system? Are there other powerful principles about memory which we could we build into the system, apart from spaced repetition?

In fact, there are ideas about memory very different from spaced repetition, but of comparable power. One such idea is *elaborative encoding*. Roughly speaking, this is the idea that the richer the associations we have to a concept, the better we will remember it. As a consequence, we can improve our memory by enriching that network of associations.

This is in some sense an obvious idea, according well with everyday experience. For instance, it's part of the reason it's so much easier to learn new facts in an area we're already expert in – we quickly form associations to our existing knowledge. But just because the idea is obvious, that doesn't mean it's particularly well supported by existing media forms. There's a lot of low-hanging fruit which we can actively support inside the mnemonic medium. Indeed, several of the suggestions above already implicitly build on the idea of elaborative encoding – principles like "avoid orphan cards" are based on this. Here's three more suggestions which build on elaborative encoding:

- **Provide questions and answers in multiple forms:**

In 1971, the psychologist Allan Paivio proposed the dual-coding theory, namely, the assertion that verbal and non-verbal information are stored separately in long-term memory. Paivio and others investigated the *picture superiority effect*, demonstrating that pictures and words together are often recalled substantially better than words alone. This suggests, for instance, that the question "Who was George Washington's Vice President?" may have a higher recall rate if accompanied by a picture of Washington, or if the answer (John Adams) is accompanied by a picture of Adams. For memory systems the dual-coding theory and picture superiority effect suggest many questions and ideas. How much benefit is there in presenting questions and answer in multiple forms? Perhaps even with multiple pictures, or in audio or video (perhaps with multiple speakers of different genders, different accents, etc), or in computer code? Perhaps in a form

that demands some form of interaction? And in each case: what works best?

- **Vary the context:** In 1978, the psychologists Steven Smith, Arthur Glenberg, and Robert Bjork reported several

Steven M. Smith, Arthur Glenberg, and Robert A. Bjork, [Environmental context and human memory](#) (1978).

experiments studying the effect of place on human memory. In one of their experiments, they found that studying material in two different places, instead of twice in the same place, provided a 40% improvement in later recall. This is part of a broader pattern of experiments showing that varying the context of review promotes memory. We can use memory systems to support things like: changing the location of review; changing the time of day of review; changing the background sound, or lack thereof, while reviewing. In each case, experiments have been done suggesting an impact on recall. It's not necessarily clear how robust the results are, or how reproducible – it's possible some (or all) are the results of other effects, uncontrolled in the original experiment. Still, it seems worth building systems to test and (if possible) improve on these results.

- **How do the cards interact with one another?**

What is the ideal network structure of knowledge? This is a very complicated and somewhat subtle set of questions. Let's give a simple example to illustrate the idea. We've presented the cards in the mnemonic medium as though they are standalone entities. But there are connections between the cards. Suppose you have cards: "Who was George Washington's Vice President?" (Answer: "John Adams", with a picture of Adams); "What did John Adams look like?" (Answer: a picture of Adams); perhaps a question involving a sketch of Adams and Washington together at some key moment; and so on. Now, this set of cards forms a network of interrelated cards. And you can use a memory system like *Quantum Country* to study that network. What happens to people's observed recall if you remove a card? Are there crucial lynchpin cards? Are there particularly effective network structures? Particularly effective types of relationship between cards? Crucially: are there general principles we can identify about finding the deepest, most powerful ways of representing knowledge in this system?

By now it's obvious that the prototype mnemonic medium we've developed is the tip of a much larger iceberg. What's more, the suggestions we've made and questions we've

asked here are also merely a beginning, to give you the flavor of what is possible.

## Two cheers for mnemonic techniques

When we discuss memory systems with people, many immediately respond that we should look into mnemonic techniques. This is an approach to memory systems very different to *Quantum Country*, Duolingo, Anki, and the other systems we've discussed. You're perhaps familiar with simple mnemonic techniques from school. One common form is tricks such as remembering the colors of the rainbow as the name Roy G. Biv (red, orange, yellow, green, etc). Or remembering the periodic table of elements using a [song](#).

A more complex variation is visualization techniques such as the *method of loci*. Suppose

you want to remember your shopping list. To do so using the method of loci, you visualize yourself in some familiar location – say, your childhood home. And then you visualize yourself walking from room to room, placing an item from your shopping list prominently in each room. When you go shopping, you can recall the list by imagining yourself walking through the house – your so-called *memory palace* – and looking at the items in each room.

If you've never used memory palaces this sounds like it couldn't possibly work. But even novices are often shocked by how well such techniques work, with just a small amount of practice. Experts who work hard developing these techniques can do remarkable things, like memorizing the order of a shuffled deck of cards, or lists of hundreds of digits. It's a way of using people's immensely powerful visual and spatial memories as a form of leverage for other types of memory.

Given all this, it's perhaps not surprising that we often meet people who tell us that mnemonic techniques are a much more promising approach to memory than ideas such as spaced repetition.

We're enthusiastic about such mnemonic techniques. But it's important to understand their limitations, and not be bedazzled by the impressiveness of someone who can rapidly memorize a deck of cards.

An enjoyable extended introduction to such techniques may be found in Joshua Foer's book "Moonwalking with Einstein" (2012).

A small minority of the population does not possess a mind's eye, and so cannot mentally visualize. This condition is known as aphantasia. One of us [asked on Twitter](#) if any aphantasics had tried using the method of loci, and if so how well it worked for them. The replies were remarkably heterogeneous (and striking), but most said such mnemonic techniques did not work for them. This deserves further study.

One caution concerns the range of what can be memorized using mnemonic techniques. In practice they're often quite specialized. Mnemonic experts will, for instance, use somewhat different approaches to memorize lists of digits versus decks of cards. Those approaches must be mastered separately – a heavy time investment for two narrow kinds of memory. Furthermore, the mnemonic techniques tend to be much better suited for concrete objects than abstract conceptual knowledge – it's difficult to store, say, the main points in the Treaty of Versailles in your memory palace. This doesn't mean it can't be done – mnemonic experts have developed clever techniques for converting abstract conceptual knowledge into concrete objects which can be stored in a memory palace. But, in general, an advantage of spaced repetition is that it works across a far broader range of knowledge than do any of the mnemonic techniques.

A second caution relates to elaborative encoding. The mnemonic techniques are, as you have likely realized, an example of elaborative encoding in action, connecting the things we want to memorize (say, our shopping list) to something which already has meaning for us (say, our memory palace). By contrast, when an expert learns new information in their field, they don't make up artificial connections to their memory palace. Instead, they find meaningful connections to what they already know. Those connections are themselves useful expertise; they're building out a dense network of understanding. It's a deeper and more desirable kind of expertise, connections native to the subject itself, not artificially constructed mnemonics.

All this makes us seem negative about mnemonic techniques. In fact, we're enthusiastic, and have to date certainly underused them in the mnemonic medium. What we've written here is merely meant to temper the over-enthusiasm we sometimes encounter. We've had people go so far as to tell us that mnemonics make memory a solved problem. That is simply false. But with their limitations understood, they're a powerful tool. This is particularly true for knowledge which has an arbitrary, *ad hoc* structure. For example, it's difficult to remember the colors of the rainbow because those colors are not obviously connected to anything else, unless you happen to have the [spectrum of visible light](#) memorized for other reasons! That makes a mnemonic like Roy G. Biv extremely helpful. And so mnemonic techniques should be thought of as a useful tool to use in building powerful memory systems, especially when combined with ideas such as spaced repetition.

## How important is memory, anyway?

People tend to fall into two buckets when told of the mnemonic medium. One group is fascinated by the idea, and wants to try it out. The second group is skeptical or even repulsed. In caricature, they say: “Why should I care about memory? I want deeper kinds of understanding! Can’t I just look stuff up on the internet? I want creativity! I want conceptual understanding! I want to know how to solve important problems! Only dull, detail-obsessed grinds focus on rote memory.”

It’s worth thinking hard about such objections. To develop the best possible memory system we need to understand and address the underlying concerns. In part, this means digging down far enough to identify the mistaken or superficial parts of these concerns. It also means distilling as sharply as possible the truth in the concerns. Doing both will help us improve and go beyond the current prototype mnemonic medium.

One response to such objections is the argument from lived experience. In the past, one of us (MN) has often helped students learn technical subjects such as quantum mechanics. He noticed that people often think they’re getting stuck on esoteric, complex issues. But, as suggested in the introduction to this essay, often what’s really going on is that they’re having a hard time with basic notation and terminology. It’s difficult to understand quantum mechanics when you’re unclear about every third word or piece of notation. Every sentence is a struggle.

It’s like they’re trying to compose a beautiful sonnet in French, but only know 200 words of French. They’re frustrated and think the trouble is the difficulty of finding a good theme, striking sentiments and images, and so on. But really the issue is that they have only 200 words with which to compose.

At the time, MN’s somewhat self-satisfied belief was that if people only focused more on remembering the basics, and worried less about the “difficult” high-level issues, they’d find the high-level issues took care of themselves. What he didn’t realize is that this also applied to him. When he began using the memory system Anki to read papers in new fields, he found it almost unsettling how much easier Anki made learning the basics of such subjects. And it made him start wondering if memory was often a binding constraint in learning new fields.

One particularly common negative response to the mnemonic medium is that people don’t want to remember “unimportant details”, and are

just looking for “a broad, conceptual understanding”. It’s difficult to know what to make of this argument. Bluntly, it seems likely that such people are fooling themselves, confusing a sense of enjoyment with any sort of durable understanding.

Imagine meeting a person who told you they “had a broad conceptual understanding” of how to speak French, but it turned out they didn’t know the meaning of “bonjour”, “au revoir”, or “tres bien”. You’d think their claim to have a broad conceptual understanding of French was hilarious. If you want to understand a subject in any real sense you need to know the details of the fundamentals. What’s more, that means not just knowing them immediately after reading. It means internalizing them for the long term.

A better model is that conceptual mastery is actually enabled by a mastery of details. One user of *Quantum Country* told

us that she found the experience of reading unexpectedly relaxing, because she “no longer had to worry” about whether she would remember the content. She simply trusted that the medium itself would ensure that she did. And she reported that she was instead able to spend more of her time on conceptual issues.

When people respond to the mnemonic medium with “why do you focus on all that boring memory stuff?”, they are missing the point. By largely automating away the problem of memory, the mnemonic medium makes it easier for people to spend more time focusing on other parts of learning, such as conceptual issues.

Another common argument against spaced repetition systems is that it’s better to rely on natural repetition. For instance, if you’re learning a programming language, the argument goes, you shouldn’t memorize every detail of that language. Instead, as you use the language in real projects you’ll naturally repeatedly use, and eventually commit to memory, those parts of the language most important to learn.

There are important partial truths in this. It is good to use what you’re learning as part of your creative projects. Indeed, an ideal memory system might help that happen, prompting you as you work, rather than in an artificial card-based environment. Furthermore, a common failure mode with memory systems is that people attempt to memorize things they’re unlikely to ever have any use for. For instance, it’s no good (but surprisingly common) for someone to memorize lots of details of a programming language they plan to use for just one small project. Or to

The last two paragraphs are adapted from our forthcoming mnemonic essay: Andy Matuschak and Michael Nielsen, [How quantum teleportation works](#) (2019).

See [here](#) and [here](#) for more on learning new fields using Anki. The last four paragraphs are adapted from [Augmenting Long-term Memory](#) (2018).

memorize details “just in case” they ever need them. These patterns are mistakes.

But the truths of the last paragraph also have limits. If you’re learning French, but don’t know any French speakers, then waiting for “natural opportunities” to speak just won’t work. And even if you do have (or create) opportunities to speak, it’s desirable to accelerate the awkward, uncomfortable early stages that form such a barrier to using the language.

It’s in this phase that memory systems shine. They can accelerate people through the awkward early stages of learning a subject. Ideally, they’ll support and enable work on creative projects. For this to work well takes good heuristics for what any given person should commit to memory; what is good for one person to memorize may be bad for another. Working such heuristics out is an ongoing challenge in the design of memory systems.

(Incidentally, a surprising number of people say they are “repulsed”, or some similarly strong word, by spaced-repetition memory systems. Their line of argument is usually some variant on: it is claimed that spaced-repetition systems help with memory; if that is true I *must* use the systems; but I hate using the systems. The response is to deny the first step of the argument. Of course, the mistake is elsewhere: there is absolutely no reason anyone “should” use such systems, even if they help with memory. Someone who hates using them should simply choose not to do so. Using memory systems is not a moral imperative!)

An immense amount of research has been done on the relationship of memory to mastery. Much of this research is detailed and context specific. But at the level of broader conclusions, one especially interesting series of studies was done in the 1970s by Herbert Simon and his collaborators. They studied chess players, and discovered that when master chess players look at a position in chess they don’t see it in terms of the individual pieces, a rook here, a pawn there.

Instead, over years of playing and analyzing games the players learn to recognize somewhere between 25,000 and 100,000 patterns of chess pieces. These much more elaborate “chunks” are combinations of pieces that the players perceive as a unity, and are able to reason about at a higher level of abstraction than the individual pieces. At least in part it’s the ability to recognize and reason about these chunks which made their gameplay so much better than novices. Furthermore, although Simon did this work

See, e.g., William G. Chase and Herbert A. Simon, *Perception in Chess* (1973). Some fascinating earlier work in a related vein was done by Adrian D. de Groot, and summarized in his book *Thought and Choice in Chess* (1965).

in the context of chess, subsequent studies have found [similar results](#) in other areas of expertise. It seems plausible, though needs further study, that the mnemonic medium can help speed up the acquisition of such chunks, and so the acquisition of mastery.

So, does all this mean we’re fans of rote memory, the kind of forced memorization common in schools?

Of course not. What we do believe is that many people’s dislike of rote memorization has led them to a generalized dislike of memory, and consequently to underrate the role it plays in cognition. Memory is, in fact, a central part of cognition. But the right response to this is not immense amounts of dreary rote memorization. Rather, it’s to use good tools and good judgment to memorize what truly matters.

We’ve identified some ways in which criticisms of memory systems are mistaken or miss the point. But what about the ways in which those criticisms are insightful? What are the shortcomings of memory systems? In what ways should we be wary of them?

We’ve already implicitly mentioned a few points in this vein. Think about problems like the need to avoid orphan questions. Or to make sure that users don’t merely learn surface features of questions. These are ways in which memory systems can fail, if used poorly. Here’s a few more key concerns about memory systems:

- **Memory systems don’t make it easy to decide what to memorize:** Most obviously, we meet a lot of people who use memory systems for poorly chosen purposes. The following is surprisingly close to a transcript of a conversation we’ve both had many times:

“I don’t like [memory system]. I tried to memorize the countries in Africa, and it was boring.”  
“Why were you trying to remember the countries in Africa?”  
[blank look of confusion.]

It’s easy to poke fun at this kind of thing. But we’ve both done the equivalent in our own memory practices. Even some users of *Quantum Country* seem to be going through the motions out of some misplaced sense of

We’ve met many mathematicians and physicists who say that one reason they went into mathematics or physics is because they hated the rote memorization common in many subjects, and preferred subjects where it is possible to derive everything from scratch. But in conversation it quickly becomes evident that they have memorized an enormous number of concepts, connections, and facts in their discipline. It’s fascinating these people are so blind to the central role memory plays in their own thinking.

duty. The question “what will be beneficial to memorize” is fundamental, and answering that question well is not trivial.

- **What’s the real impact of the mnemonic medium on people’s cognition?** How does it change people’s behavior? A famous boxer is supposed to have said that everyone has a plan until they get punched in the face. Regular users of memory systems sometimes report that while they can remember answers when being tested by their system, that doesn’t mean they can recall them when they really need them. There can be a tip-of-the-tongue feeling of “Oh, I know this”, but not actual recall, much less the fluent facility one ultimately wants for effective action. Furthermore, the user may not even recognize opportunities to use what they have learned. More broadly: memory is not an end-goal in itself. It’s embedded in a larger context: things like creative problem-solving, problem-finding, and all the many ways there are of taking action in the world. We suspect the impact of memory systems will vary a lot, depending on their design. They may be used as crutches for people to lean on. Or they may be used to greatly enable people to develop other parts of their cognition. We don’t yet understand very well how to ensure they’re enablers, rather than crutches. But later in the essay we’ll describe some other tools for thought that, when integrated with memory systems, may better enable this transition to more effective action.

## How to invent Hindu-Arabic numerals?

Let’s briefly get away from memory systems. Imagine you’re a designer living in ancient Rome, working for MDC (Mathematical Designs Corporation). A client comes in one day, expressing a desire to improve on Roman numerals. Of course, that’s not literally how they describe their problem to you – more likely it’s a tax collector wanting to tabulate taxes more efficiently, and having some vague notion that MDC may be able to help. But to you, an experienced designer, it seems that an improved system of numerals may be what they need.

How should you respond to this request? From our modern vantage point we know a vastly better system of numerals is possible, the Hindu-Arabic numerals. Hindu-Arabic numerals were, in fact, a great leap in the history of tools for thought. Could you, as a designer, have made that leap? What creative steps would be needed to invent Hindu-Arabic numerals, starting from the Roman

numerals? Is there a creative practice in which such steps would be likely to occur?

To be clear: this is a somewhat fanciful thought experiment. Many of the ideas needed to get to Hindu-Arabic numerals were, in fact, known earlier to the Babylonians, to the Greeks, and in other cultures. They were also inchoate in the abacus and similar devices. And so we’re not asking a literal historical question. Rather, it’s a question meant to stimulate thought: what *design process* could take you from Roman numerals to Hindu-Arabic numerals?

We can’t know the answer to this question for sure. But it’s worth pointing out that the Hindu-Arabic numerals aren’t just an extraordinary piece of design. They’re also an extraordinary mathematical insight. They involve many non-obvious ideas, if all you know is Roman numerals. Perhaps most remarkably, the meaning of a numeral actually *changes*, depending on its position within a number. Also remarkable, consider that when we add the numbers 72 and 83 we at some point will likely use  $2+3=5$ ; similarly, when we add 27 and 38 we will also use  $2+3=5$ , despite the fact that the meaning of 2 and 3 in the second sum is completely different than in the first sum. In modern user interface terms, the numerals have the same affordances, despite their meaning being very different in the two cases. We take this for granted, but this similarity in behavior is a consequence of deep facts about the number system: commutativity, associativity, and distributivity. All these properties (and many more) point to the design and mathematical insights being inextricably entangled: the mathematical insights are, in some sense, design insights, and vice versa.

The same phenomenon occurs in the conventional grade-school algorithms for multiplication and division. One of us has spun a short piece of *discovery fiction* discussing in more detail the way a hypothetical designer might have arrived at these ideas.

Indeed, it seems fair to say that any person who could invent Hindu-Arabic numerals, starting from the Roman numerals, would be both one of the great mathematical geniuses who ever lived, and one of the great design geniuses who ever lived. They’d have to be extraordinarily capable in both domains, capable of an insight-through-making loop which used the evolving system of numerals to improve not just their own mathematical ideas, but to have original, world-class insights into mathematics; and also to use those mathematical insights to improve their evolving system of numerals.

This is rather sobering if we compare to conventional modern design practice. In a typical practice, you’d interview domain experts (in this case, mathematicians),

and read any relevant literature. You'd talk to users of existing systems, and analyze serious behavior, both individually and at scale. In short, you'd do what people in the design community refer to as immersing themselves in the target field.

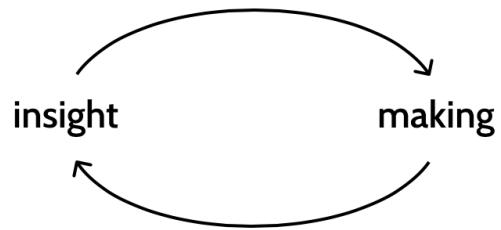
This is a powerful practice. At its best it causes systems to come into existence which would otherwise be inconceivable. If applied to Roman numerals (in hypothetical ancient Rome, not today) this practice would likely improve them a great deal. But it would not provide anywhere near the level of mathematical insight needed to arrive at Hindu-Arabic numerals.

Our story about Hindu-Arabic numerals and mathematics is fanciful. But it expresses a general truth: *the most powerful tools for thought express deep insights into the underlying subject matter*. In the case of memory systems, this means they're not just "applied cognitive science", a collage of existing ideas from cognitive science pasted together using modern design practice. Rather, they will express deep original insights into memory, insights no-one else in the world has ever had. A truly great memory system will be cognitive science of the highest order.

From this discussion, we take away a warning and an aspiration.

The warning is this: conventional tech industry product practice will not produce deep enough subject matter insights to create transformative tools for thought. Indeed, that's part of the reason there's been so little progress from the tech industry on tools for thought. This sounds like a knock on conventional product practice, but it's not. That practice has been astoundingly successful at its purpose: creating great businesses. But it's also what Alan Kay has dubbed a pop culture, not a research culture. To build transformative tools for thought we need to go beyond that pop culture.

The aspiration is for any team serious about making transformative tools for thought. It's to create a culture that combines the best parts of modern product practice with the best parts of the (very different) modern research culture. You need the insight-through-making loop to operate, whereby deep, original insights about the subject feed back to change and improve the system, and changes to the system result in deep, original insights about the subject.



Note that we are not making the common argument that making new tools can lead to new subject matter insights for the toolmaker, and vice versa. This is correct, but is much weaker than what we are saying. Rather: making new tools can lead to new subject matter insights *for humanity as a whole* (i.e., significant original research insights), and vice versa, and this would ideally be a rapidly-turning loop to develop the most transformative tools.

Doing this is a cultural struggle. It seems to be extraordinarily rare to find the insight-through-making loop working at full throttle. People with expertise on one side of the loop often have trouble perceiving (much less understanding and participating in) the nature of the work that goes on on the other side of the loop. You have researchers, brilliant in their domain, who think of making as something essentially trivial, "just a matter of implementation". And you have makers who don't understand research at all, who see it as merely a rather slow and dysfunctional (and unprofitable) making process. This is certainly true in Silicon Valley, where it's common to meet accomplished technical makers who, after reading a few stories from Richard Hamming and Richard Feynman, think they understand research well enough that they can "create the new Bell Labs". Usually they're victims of Dunning-Krugeritis, so ignorant they're not even aware of their ignorance.

Of course, we've got a long way to go with *Quantum Country*. It's not yet generating nearly deep enough ideas about memory and cognition; it's not yet one of the world's foremost memory laboratories. And considered as a product, it's also in the very earliest days; we're not yet iterating nearly fast enough, nor learning nearly fast enough from the system. Getting the insight-through-making loop to operate at full throttle will mean reinventing parts of both research culture and conventional product development culture; it will mean new norms and a new type of person involved in key decision making. But that's the aspiration, and what we believe is necessary to develop transformative tools for thought.

## Part II: Exploring tools for thought more broadly

We've examined the mnemonic medium in some depth. The intent was to show you the early stages in the development of a specific tool for thought, and some of the thinking enabled by that development. In this second part of the essay we explore more broadly, briefly sketching ideas for several other tools for thought. And we'll address some broader questions, especially around why there hasn't been more work on tools for thought.

### Mnemonic video

In 2014, the digital artist Eric Wernquist released an extraordinary short video entitled "[Wanderers](#)". The video provides a first-person glimpse of what it would be like to explore the Solar System:



We hear the narrator's (Carl Sagan) wonder and awe, and cannot help but empathize with his deep belief in the value of exploration. We get a sense of how many mysteries and how much beauty there is in our own cosmic neighborhood. The music begins with a wistful nostalgia for those of our ancestors who dared to explore, and then changes to convey excitement and danger and the boldness of those members of our and future generations who continue that exploration.

It's interesting to contrast the video to the video's script, a short text by Carl Sagan. The [text](#) is beautiful, but reading it is a much more remote and cerebral experience, conveying a much less visceral emotional understanding.

We have a friend, Grant Sanderson, who makes astonishing mathematics videos on his YouTube channel, [3Blue1Brown](#). One of our favorites is "[Who cares about topology? \(Inscribed rectangle problem\)](#)", a video sketching a proof of a relatively recent research result in geometry, using ideas from algebraic topology. This sounds fearsome, but the video is beautiful and accessible, and has been viewed more than 1.2 million times.

As with *Wanderers*, watching this video is a remarkable emotional experience. It's obvious the person narrating the video loves mathematics, and you cannot help but empathize. As you watch, you experience repeated "Ahah!" moments, moments of surprising insight, as connections that were formerly invisible become obvious. It shows mathematics as something beautiful, containing extraordinary ideas and intriguing mysteries, while at the same time showing that doing mathematics is not mysterious, that it is something anyone can understand and even do.

It's tempting to overlook or undervalue this kind of emotional connection to a subject. But it's the foundation of all effective learning and of all effective action. And it is much easier to create such an emotional connection using video than using text.

There's a flipside to this emotional connection, however. We've often heard people describe Sanderson's videos as about "teaching mathematics". But in conversation he's told us he doesn't think more than a small fraction of viewers are taking away much detailed understanding of mathematics. We suspect this is generally true, that high affect videos usually do little to change people's detailed intellectual understanding. Rather, the extraordinary value of the videos lies in the emotional connection they create.

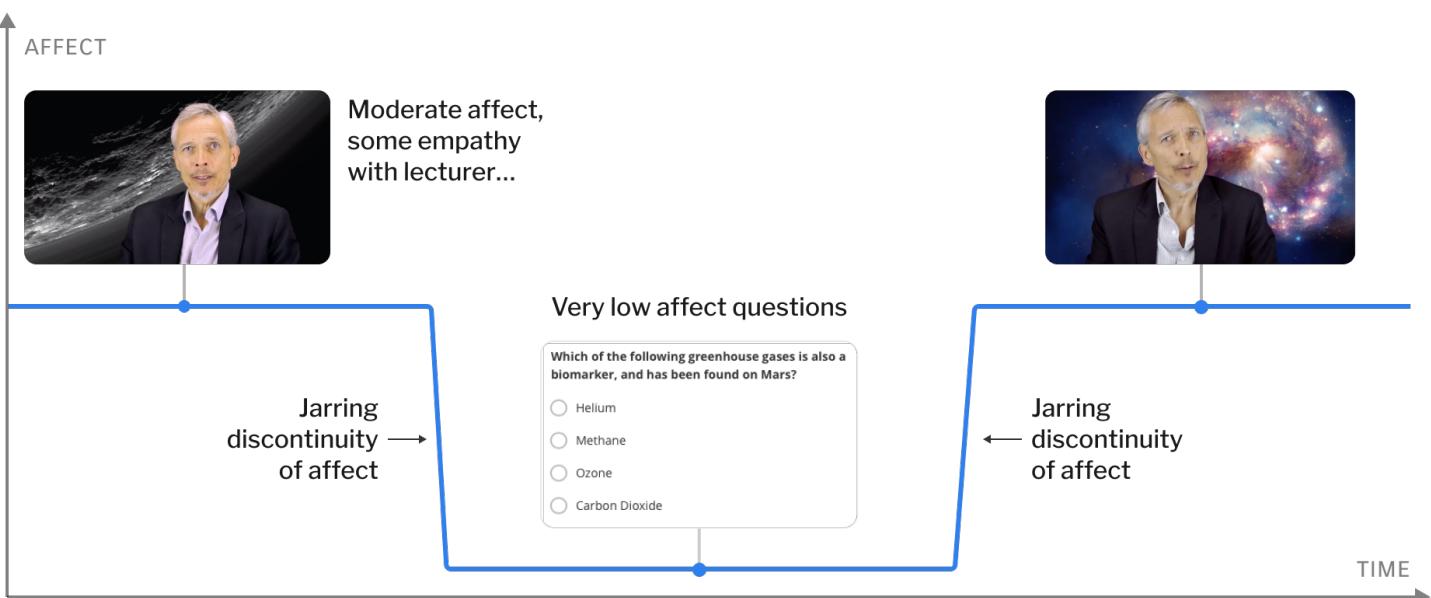
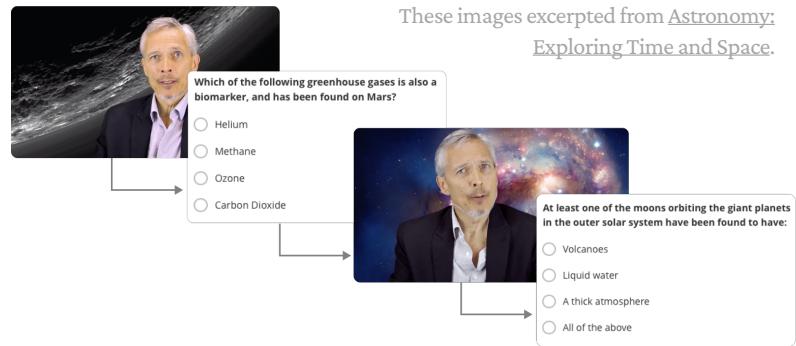
Is it possible to create a medium which blends the best qualities of both video and text?

In particular, is it possible to create a medium which has the emotional range possible in video – a range which can be used to convey awe and mystery and surprise and beauty? But which can also firmly ground that emotional connection in detailed understanding, the mastery of details which is the *raison d'être* of both conventional text and, perhaps even moreso, of the mnemonic medium?

We believe this may be possible, and we plan to develop a *mnemonic video* form that provides both the emotional connection possible in video, and the mastery of details possible in the mnemonic medium.

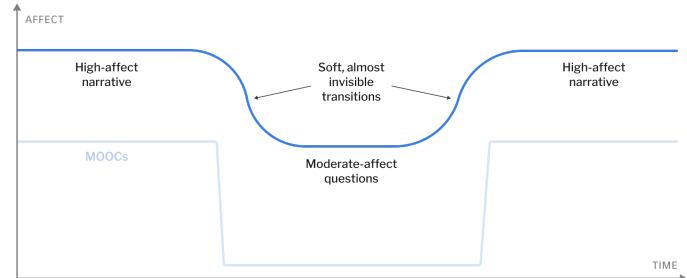
Creating such a form is challenging. Many MOOC platforms have attempted something a little in this vein. The typical approach is to have a low-affect talking head video, with the videos interrupted occasionally for brief quizzes. At right, we depict how it works on one MOOC platform, Coursera.

Other MOOCs differ in the details. But the overall emotional experience may be summed up in the plot below:



The very best parts of the video may be emotionally compelling, though it's rare that they achieve the emotional range and connection of the best videos from people like Grant Sanderson. And the overall emotional experience is disjointed, almost repellent. Is it possible to create an integrated medium, with a unified and carefully crafted emotional and intellectual experience? Ideally, is it possible to create something like the plot at right?

In MOOCs, questions are typically presented in a very dry form, detached from context. In the mnemonic video the narrator would explain why the questions are important, and why the user will benefit from participating, as a seamless part of the overall narration. Done right – perhaps with appropriate music, and a sense of urgency or play in the narration – it would create a real sense of the stakes. At the same time, the video player can be modified so the user can respond directly to questions, as part of the spaced-repetition experience. The result would be much softer transitions between the high-affect core narration and the moderate-affect questions.



Here's a short sketch of one approach to doing this, showing how the narrator could ask questions aloud as an integrated part of the overall narration:

The narrator's script in this sketch is adapted from *Cosmos* (1984), episode 4.

1.

NARRATOR:

In 1908, a piece of a comet hit the Earth. Hurtling at more than 100,000 km/h, it was a mountain of ice the size of a football field and weighing almost a million tons.



Dramatic visuals and narrative pathos establish high affect

4.

NARRATOR:

It's because the ices in the comet are all melted in the impact. There's going to be very few recognizable pieces of comet left on the ground.



The narrator answers the question inline as the video continues

2.

NARRATOR:

A cometary fragment will produce a great radiant fireball and a mighty blast wave. It'll burn trees and level forests. Yet despite all that destruction, can you see why comets usually leave no crater in the ground?

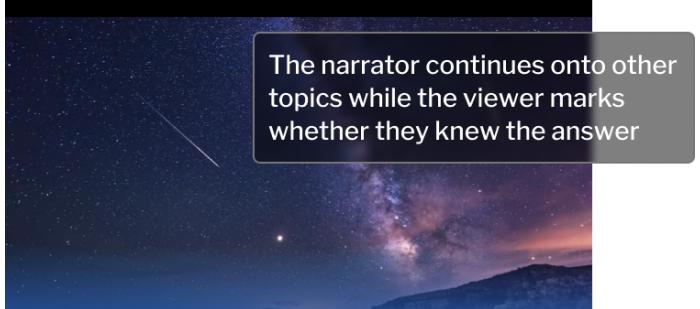


The narrator smoothly pivots to a question for the viewer, right in the middle of the video.

5.

NARRATOR:

We humans like to think of the heavens as stable, serene, unchanging. But comets suddenly appear and hang ominously in the sky, night after night, for weeks



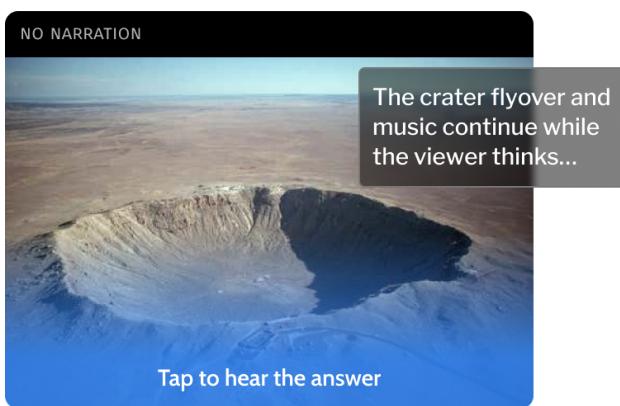
The narrator continues onto other topics while the viewer marks whether they knew the answer

3.

NO NARRATION



Did you get that answer?



The crater flyover and music continue while the viewer thinks...

Tap to hear the answer

The question interface shares the screen with the video

We keep friction low to maintain immersion: the input is "tap anywhere."

6.

NARRATOR:

So the idea developed that the comet had to be there for a reason. The reason was that comets were predictions of disaster, that they foretold the fall of kingdoms."



The interface recedes completely after the viewer marks their answer

It seems likely that the rhythm of mnemonic videos would be quite different to mnemonic essays. In particular, the frequency and density of questions would be lower than in the mnemonic medium, and it would be necessary to test different beats and cadences to ensure a good balance of emotional and intellectual experience. Even high-affect video typically has quieter moments; it achieves the high affect in part by contrast to the lower-intensity moments. Think about the way a good action movie or thriller needs lulls; if it's too high-intensity all the time, eventually our emotional response is dulled. We could design mnemonic video so the questions help fill this lower-intensity emotional beat.

Of course, this is merely a quick sketch of one approach to the design of mnemonic video. Ideally, there would also be a spaced-repetition component, perhaps with the questions asked in text instead of video. This sort of sketch seems to us a promising direction, but needs considerable development and intense testing. In particular, we need to do detailed, second-by-second user experience testing, to understand and shape users' emotional and intellectual experience. That would continue until we were confident that our target users were having the desired experience. Ideally, we'd also generate several more very different designs, and try to understand how each approach would impact the user's emotional and intellectual experience.

The broader point here is about taking emotion seriously. Historically, a lot of work on tools for thought has either ignored emotion, or treated it as no more than a secondary concern. Instead, that work has focused on new skills acquired, on what the user "learns". They've been designing for Spock, when emotional connection is a high-order bit. Do users feel disinterested? Afraid? Hostile? Anxious? Or do they internalize a sense of excitement, of beauty, perhaps even an expansion in their own goals, an expansion of their self?

By contrast, media forms such as movies and music and (often) video games do take emotion seriously. The designers of such forms often have incredibly elaborate models of user's emotional responses. These models range from detailed, second-by-second understanding, to deep thinking about a user's overall emotional journey. We believe it's possible and desirable to use such approaches in the development of tools for thought.

At the same time, a positive emotional experience alone is not enough. For tools for thought to attain enduring power, the user must experience a real growth in mastery, an expansion in their ability to act. And so we'd like to take both the emotional and intellectual experience of tools for

thought seriously. Mnemonic video is a good venue for such exploration. To paraphrase Einstein, attaining a detailed understanding without forming an emotional connection is lame; while forming an emotional connection without detailed understanding has no enduring power.

## Why isn't there more work on tools for thought today?

If tools for thought are so great, why isn't more work being done on them? Why aren't they a major industry?

As noted in the introduction, there's certainly a lot of lip service paid. It is, for instance, common to hear technologists allude to Steve Jobs's metaphor of computers as "bicycles for the mind". But in practice it's rarely more than lip service. Many pioneers of computing have been deeply disappointed in the limited use of computers as tools to improve human cognition. Douglas Engelbart [disparaged](#) the "dangerous, disappointing, narrow, path that we seem to be stuck with following". When asked in 2006 how much of his vision had been achieved, Engelbart replied facetiously "[about 2.8 percent](#)". Alan Kay gives talks asserting "[The real computer revolution hasn't happened yet](#)" and in [an interview](#) has described the modern web as "reinventing the flat tire... at least give us what Engelbart did, for Christ's sake."

Our experience is that many of today's technology leaders genuinely venerate Engelbart, Kay, and their colleagues. Many even feel that computers have huge potential as tools for improving human thinking. But they don't see how to build good businesses around developing new tools for thought. And without such business opportunities, work languishes.

What makes it difficult to build companies that develop tools for thought? To answer this, consider Adobe, one of the few large companies serious about developing new tools for thought. It's poured money into developing new mediums for designers and artists – programs such as *Illustrator*, *Photoshop*, and so on. These mediums are remarkable tools for thought.

Unfortunately for Adobe, such mediums are extremely expensive to develop, and it's difficult to prevent other companies from cheaply copying the ideas or developing near-equivalents. Consider, for example, the way the program *Sketch* has eaten into

Another large company which takes tools for thought extremely seriously is AutoDesl. A similar story could be told about it.

The plural of medium is, of course, media. However, in this context media would usually mean many pieces of new content. That's not what we mean: we mean multiple different new mediums (*Illustrator*, *Photoshop* etc). We'll reserve the unusual pluralization for this somewhat unusual meaning.

Adobe's market share, after duplicating many of the best features from several of Adobe's products, perhaps most notably *Illustrator*. And consider the way *Figma* is now eating into both *Sketch* and *Illustrator*'s market share. Both *Sketch* and *Figma* have done this without needing to make an enormous investment in research. That's a big advantage they have over Adobe.

As Marc Andreessen has observed:

In Elad Gil's "High Growth Handbook" (2018).

true defensibility purely at the product level is really rare in [Silicon] Valley, because there are a lot of really good engineers... And then there's the issue of leap-frogging. The next team has the opportunity to learn from what you did and then build something better.

Put another way, many tools for thought are public goods. They often cost a lot to develop initially, but it's easy for others to duplicate and improve on them, free riding on the initial investment. While such duplication and improvement is good for our society as a whole, it's bad for the companies that make that initial investment. And so such tools for thought suffer the fate of many public goods: our society collectively underinvests in them, relative to the benefits they provide.

Earlier, we argued that modern design practice generally isn't up to the challenge of producing genuinely transformative tools for thought. On the surface, that process-level argument appears very different to the public goods argument we just made. In fact, the process-level explanation is a consequence of the public goods explanation: companies don't use the necessary processes because there's little value to them in doing so. By contrast, in "harder-tech" industries – say, chip design – companies have much more incentive to do deep research work. In those industries it's considerably harder for other companies to duplicate or capture the value of that research.

It's illuminating to contrast with video games. Game companies develop many genuinely new interface ideas. This perhaps seems surprising, since you'd expect such interface ideas to also suffer from the public goods problem: game designers need to invest enormous effort to develop those interface ideas, and they are often immediately copied (and improved on) by other companies, at little cost. In that sense, they are public goods, and enrich the entire video game ecosystem.

But there's a big difference between video game companies and companies such as Adobe. Many video games make most of their money from the first few months of sales. While other companies can (and do) come in and copy or riff on any new ideas, it often does little to affect revenue from the original game, which has already made most of its money. While this copying is no doubt irritating for the companies being copied, it's still worth it for them to make the up-front investment.

The net result is that in gaming, clever new interface ideas can be distinguishing features which become a game's primary advantage in the marketplace. Indeed, new interface ideas may even help games become classics – consider the many original (at the time) ideas in games ranging from *Space Invaders* to *Wolfenstein 3D* to *Braid* to *Monument Valley*. As a result, rather than under-investing, many companies make sizeable investments in developing new interface ideas, even though they then become public goods. In this way the video game industry has largely solved the public goods problems.

By contrast, a company like Adobe builds their business around distribution and long-term lock in. They convince people – indeed, entire organizations – to make long-term commitments to their products. Schools offer classes so people can call themselves "*Photoshop experts*" or "*Illustrator experts*". Companies designate their design departments as "*Adobe shops*". So while Adobe does invest in developing clever new interface ideas (for them, unlike the video game companies, this genuinely means tools for thought), it's less central to their competitive advantage, and they invest less than they would if it was their central advantage. And Adobe does perhaps as much or more work developing tools for thought as any company.

It's encouraging that the video game industry can make inroads on the public goods problem. Is there a solution for tools for thought? Unfortunately, the novelty-based short-term revenue approach of the game industry doesn't work. You want people to really master the best new tools for thought, developing virtuoso skill, not spend a few dozen

In fact, cloning is a real issue in gaming, especially in very technically simple games. An example is the game *Threes*, which took the developers more than a year to make. Much of that time was spent developing beautiful new interface ideas. The resulting game was so simple that clones and near-clones began appearing within days. One near clone, a game called *2048*, sparked a mini-craze, and became far more successful than *Threes*. At the other extreme, some game companies prolong the revenue-generating lifetime of their games with re-releases, long-lived online versions, and so on. This is particularly common for capital-intensive AAA games, such as the *Grand Theft Auto* series. In such cases the business model relies less on clever new ideas, and more on improved artwork (for re-release), network effects (for online versions), and branding.

hours (as with most games) getting pretty good, and then moving onto something new.

Another plausible solution to the public goods problem is patents, granting a temporary monopoly over use of an invention. Many software companies, including Adobe, develop a large patent portfolio. However, the current patent system is not a solution for this problem. In 2017, Dana Rao, Adobe's Vice President for Intellectual Property and Litigation, posted a [call](#) for major reforms to the patent system, stating that:

[the patent] system is broken... What happened? A patent gold rush built by patent profiteers... Their value lies not in the innovation behind the patent but in the vagueness of the patent's claims and the ability to enforce it in a plaintiff-friendly forum... Where did the material for these bad patents come from? The advent of software... This led to idea-only patents being granted with broad and often invalid claims, and eager patent profiteers were only too glad to take advantage.

Adobe shares in common with many other software companies that much of their patenting is defensive: they patent ideas so patent trolls cannot sue them for similar ideas. The situation is almost exactly the reverse of what you'd like. Innovative companies can easily be attacked by patent trolls who have made broad and often rather vague claims in a huge portfolio of patents, none of which they've worked out in much detail. But when the innovative companies develop (at much greater cost) and ship a genuinely good new idea, others can often copy the essential core of that idea, while varying it enough to plausibly evade any patent. The patent system is not protecting the right things.

Switching away from the viewpoint of individual companies, and to the viewpoint of society as a whole, not only do we want to incentivize invention, we also want ideas to move reasonably rapidly into the public domain. Think about fundamental tools for thought such as writing and the number system. Obviously, it's good that those spread throughout society, unencumbered by IP concerns! More broadly, many tools of thought become more valuable for society as they become more ubiquitous. Again, here the modern patent system has numerous well-known problems, striking a poor balance between private and public interest. While a well-designed patent system might very well help solve the public goods problem, the patent system we actually have seems poorly adapted to the problem.

Is it possible to avoid the public goods problem altogether? Here's three classes of tools for thought which do:

- Search engines such as Google are tools for thought. They avoid the public goods problem because their value is in their brand and in hard-to-duplicate and capital intensive backend elements (including their data centers, proprietary algorithms, ad network, and distribution), not in their interface ideas.
- A service such as Twitter can be considered a tool for collective thought. While the interface is easily copied, the company is hard to duplicate, due to network effects.
- Novel hardware devices (e.g., for VR, or the Wii remote, or for new musical instruments) can be used as the basis for new tools for thought. While hardware can be duplicated, it's often much more expensive than duplicating software. And, in any case, the advantage for such companies is often in distribution, marketing, and relationships with vendors who make products for the platform.

While these suggestions all avoid the public goods problem, they don't directly solve the public goods problem. And many promising directions – including ideas such as the mnemonic medium and mnemonic videos – involve a substantial public goods element. Is it possible to solve the public goods problem in such cases? The two most promising approaches seem to us to be:

- Philanthropic funding for research. This approach was used, for instance, by the field of computer animation and animated movies. Decades of public research work on computer animation resulted in a large number of powerful and (in many cases) publicly available ideas. This, in turn, helped prepare the way for companies such as Pixar and Dreamworks, which developed many of the ideas further, and took them to scale.
- The model used by Adobe and similar companies, in which new tools for thought are a central part of the company's operations, but not the core of their competitive moat. That moat may instead be built around training, marketing, documentation, and so on.

## Questioning our basic premises

There are three important premises we've taken for granted up to now. First is the assertion that we're still in the early days, that many more transformative tools for

thought are yet to be discovered. Second is the assertion that work on tools for thought is stalled, that there's not lots of interesting work going on. And third, a kind of meta-premise, is the assertion that this kind of work is worth doing, relative to the current fashion for related ideas such as artificial general intelligence and brain-computer interfaces. In this section we discuss these premises.

#### What if the best tools for thought have already been discovered?

In other words, perhaps the 1960s and 1970s were an unrepeatable golden age, and all we can expect in the future is gradual incremental improvement, and perhaps the occasional major breakthrough, at a decreasing frequency?

There's a plausible story suggesting this is true. Tech is an enormous industry, well funded, with many bright, ambitious, talented people. Surely if there were major ideas to discover, people would do so? This argument is reinforced by the fact that, at the individual level, we meet many brilliant people who are fascinated by (and often working on) tools for thought, but who nonetheless seem to be making slow progress.

But while this story has a superficial appeal, it's misleading. Really difficult problems – problems like inventing Hindu-Arabic numerals – aren't solved by good intentions and interest alone. A major thing missing is foundational ideas powerful enough to make progress. In the earliest days of a discipline – the proto-disciplinary stage – a few extraordinary people – people like Ivan Sutherland, Doug Engelbart, Alan Kay, and Bret Victor – may be able to make progress. But it's a very bespoke, individual progress, difficult to help others become proficient in, or to scale out to a community. It's not yet really a discipline. What's needed is the development of a powerful praxis, a set of core ideas which are explicit and powerful enough that new people can rapidly assimilate them, and begin to develop their own practice. We're not yet at that stage with tools for thought. But we believe that we're not so far away either.

While that argument is helpful context, it doesn't address the core point: it doesn't mean there are a lot of new transformative tools for thought waiting to be discovered. Again: maybe the most important tools for thought have already been discovered?

We can't predict the future, so it's not possible to answer this question with certainty. But it seems to us that the human race just hasn't really tried very hard yet. When small groups of motivated people do – as in pioneering labs such as PARC, SRI, and other DARPA-inspired early efforts,

as well as modern labs such as Dynamicland – they make rapid progress. It's extremely encouraging that those efforts – tiny efforts, in the scheme of humanity's overall research effort – make such rapid progress. To us, that suggests scaling them up, becoming much more ambitious.

#### Isn't this what the tech industry does? Isn't there a lot of ongoing progress on tools for thought?

In particular, aren't there already a lot of imaginative, determined, well-funded people working on this? Isn't tech in considerable part already about developing new tools for thought?

Part of this question is caused by a confusion in terms. Obviously, many tech companies build special-purpose tools for solving specific problems. But while those may be valuable tools, they're certainly not "tools for thought" in the broad sense we're discussing – not like language or writing or, for that matter, Illustrator.

Still, there are tech companies which really do develop tools for thought. We already discussed some examples where companies have partially or totally avoided the public goods problem, tools such as: Illustrator, Google Search, Twitter, Slack, Google Docs, programmer tools, and so on. All really are significant tools for thought.

But consider our most fundamental tools for thought – language, writing, music, etc. Those are public goods. No-one owns language; to the extent that it is owned (trademarks and so on) it may actually limit the utility of language. These tools are all about introducing fundamental new mental representations and mental operations. Those aren't owned by any company, they're patterns owned by humanity.

Programmer tools are a case where the insight-through-making loop operates quite well. The avoidance of the public goods problem is enabled by a complex set of factors, including the fact that this is a case where many highly skilled researchers are simultaneously also accomplished makers, and it's either their job to produce public goods (researchers) or an enjoyable hobby. To a lesser extent the same is true in the artistic and music communities.

This argument makes it seem likely that many of the most fundamental and powerful tools for thought do suffer the public goods problem. And that means tech companies focus elsewhere; it means many imaginative and ambitious people decide to focus elsewhere; it means we haven't developed the powerful practices needed to do work in the area, and a result the field is still in a pre-disciplinary stage. The result, ultimately, is that it means the most fundamental and powerful tools for thought are undersupplied.

## Why not work on artificial general intelligence (AGI) or brain-computer interfaces (BCI) instead?

We're often asked: why don't you work on AGI or BCI instead of tools for thought? Aren't those more important and more exciting? And for AGI, in particular, many of the skills required seem related.

They certainly are important and exciting subjects. What's more, at present AGI and BCI are far more fashionable (and better funded). As a reader, you may be rolling your eyes, supposing our thinking here is pre-determined: we wouldn't be writing this essay if we didn't favor work on tools for thought. But these are questions we've wrestled hard with in deciding how to spend our own lives. One of us wrote a book about artificial intelligence before deciding to focus primarily on tools for thought; it was not a decision made lightly, and it's one he revisits from time to time. Indeed, given the ongoing excitement about AGI and BCI, it would be surprising if people working on tools for thought didn't regularly have a little voice inside their head saying "hey, shouldn't you be over there instead?" Fashion is seductive.

One striking difference is that AGI and BCI are based on relatively specific, well-defined goals. By contrast, work on tools for thought is much less clearly defined. For the most part we can't point to well-defined, long-range goals; rather, we have long-range visions and aspirations, almost evocations. The work is really about exploration of an open-ended question: how can we develop tools that change and expand the range of thoughts human beings can think?

Culturally, tech is dominated by an engineering, goal-driven mindset. It's much easier to set KPIs, evaluate OKRs, and manage deliverables, when you have a very specific end-goal in mind. And so it's perhaps not surprising that tech culture is much more sympathetic to AGI and BCI as overall programs of work.

But historically it's not the case that humanity's biggest breakthroughs have come about in this goal-driven way. The creation of language – the ur tool for thought – is perhaps the most important occurrence of humanity's existence. And although the origin of language is hotly debated and uncertain, it seems extremely unlikely to have been the result of a goal-driven process. It's amusing to try imagining some prehistoric quarterly OKRs leading to the development of language. What sort of goals could one possibly set? Perhaps a quota of new irregular verbs? It's inconceivable!

Similarly, the invention of other tools for thought – writing, the printing press, and so on – are among our

greatest ever breakthroughs. And, as far as we know, all emerged primarily out of open-ended exploration, not in a primarily goal-driven way. Even the computer itself came out of an exploration that would be regarded as ridiculously speculative and poorly-defined in tech today. Someone didn't sit down and think "I need to invent the computer"; that's not a thought they had any frame of reference for. Rather, pioneers such as Alan Turing and Alonzo Church were exploring extremely basic and fundamental (and seemingly esoteric) questions about logic, mathematics, and the nature of what is provable. Out of those explorations the idea of a computer emerged, after many years; it was a discovered concept, not a goal. Fundamental, open-ended questions seem to be at least as good a source of breakthroughs as goals, no matter how ambitious. This is difficult to imagine or convince others of in Silicon Valley's goal-driven culture. Indeed, we ourselves feel the attraction of a goal-driven culture. But empirically open-ended exploration can be just as, or more successful.

"What will new tools for thought be like?" is a question we hear often. And yet, almost by definition, we cannot say. As we noted earlier, if we could communicate the experience in an essay, then the tools would be failing at their job; they would not be transforming a person's thinking, or even their consciousness. Concretely: to understand the mnemonic medium you must use it intensively over an extended period. And even then you may not be conscious of the effect; we've done interviews with users who are apparently unaware of the incredible level of recall they have of material in the essay they have read. One of the most famous papers in the philosophy of consciousness is entitled "What is it like to be a bat?" Each tool for thought poses a similar question, near impossible to answer without immersion in the tool: "What is it like to be a language user? A musician?" and so on.

It seems plausible to us that work on tools for thought will be, over the next few decades, more important than work on AGI and BCI. And, given how fashionable and well-funded work on AGI and BCI currently is, it seems nearly certain that work on tools for thought offers vastly greater benefit, at the margin.

What about the longer term? There, the situation is less clear. It seems likely that the three fields will merge, or at least feed strongly into one another. Together with Shan Carter, one of us has argued that one of the most promising applications for AI is as a way of discovering new tools for thought.

Shan Carter and Michael Nielsen, [Using Artificial Intelligence to Augment Human Intelligence \(2017\)](#).

BCI seems likely to be even more closely related. BCI is sometimes described using ideas like a memory chip for long-term memories, or some way of increasing short-term working memory. Such ideas may well become important. But it also seems possible that BCIs will be used to enable new mental operations, new mental representations, and new affordances for thought; in short, the same kind of things as are involved in developing non-BCI tools for thought. Perhaps we'll develop the capacity to directly imagine ourselves in 4 or 5 or more dimensions; or traversing a Riemann manifold; or the ability to have multiple tracks of conscious attention. These are about changing the interface for thought, the basic abstractions and operations which are allowed. And so it seems plausible that work today on tools for thought will directly impact the way we use BCIs in the future.

## Executable books

*The skill of writing is to create a context in which other people can think.*

— Edwin Schlossberg

The computer scientist Peter Norvig has written an [interactive essay](#) discussing the distribution of wealth in society. Norvig's essay is a Jupyter notebook which expresses many of the ideas in running Python code. That code sets up a population of agents, with an initial distribution of wealth. Agents randomly (and repeatedly) meet one another in pairs, and engage in simple economic transactions. More concretely: a simple transaction model could be that when two people meet, their joint wealth is pooled, and then randomly divided between the two of them. That model is just to give you the gist – more complex transaction models are, of course, possible. The notebook simulates how the distribution of wealth evolves over time.

Part of what makes Norvig's essay beautiful is that with just a few lines of Python code Norvig is able to show some surprising results about wealth inequality. For instance, his results suggest that the initial distribution of wealth in the economy doesn't much affect the long-run distribution of wealth. Rather, it's the nature of the transactions which determines the long-run distribution of wealth. This likely violates at least some user's intuitions. As another example, his results also suggest that constraining agents to trade only with people geographically near them makes little difference to the final distribution of wealth.

Results like these will challenge the intuition of some users. But instead of those challenges being on the basis of

easily-ignored abstract arguments, users can immediately engage with Norvig's model. Suppose someone doesn't like the idea that the initial distribution of wealth doesn't affect long-run wealth inequality. They're challenged to find a counterexample, an initial distribution of wealth which does affect long-run inequality. They can experiment easily, making simple modifications to just one or a few lines of Python code, trying to find instances where the initial distribution matters. No matter whether they succeed or fail, they will build a better understanding of the problem.

Sufficiently motivated reasoners will, of course, ignore any conclusion they don't like. Such people are most likely a lost cause to serious thought.

Suppose the content of Norvig's essay had instead been presented in a more conventional static form. For a reader to extend or interrogate the results would require total mastery of the material, and a high level of mathematical competence. But in the notebook format it's much easier for the reader to experiment. Their exploration is scaffolded, they can make small modifications and see the results, even the answers to questions Norvig did not anticipate. This kind of scaffolded exploration is a way to build up their own understanding, and perhaps even push the frontiers of knowledge.

Norvig's essay is one of thousands (or perhaps even millions) of Jupyter notebooks that have been created. Of course, most such notebooks are hastily and poorly written. But in the hands of an excellent writer and thinker like Norvig, notebooks can become remarkable environments for thought, both individual and shared. It's tempting to regard them as merely a mashup of essay and code. But really they're a new media form, with different possibilities from either essays or code, and with striking opportunities to go much further. In this section we explore those opportunities.

We described Norvig's essay as an "interactive essay". It's useful to have a more specific term, to distinguish it from other interactive forms, like the mnemonic medium. In this essay, we'll use the term "executable book". We won't define this precisely here; definition is not the point. Rather, the point is to try to better understand the potential of media forms which combine prose and code in something like this form.

Of course, systems like Jupyter go back decades. There are antecedents in Knuth's notion of literate programming, in Mathematica notebooks, in PARC's Learning Research Group, in the PLATO system (and, more broadly, computer-assisted instruction), to name but a few. In the discussion below we emphasize opportunities that seem to us undervalued in many of these prior systems.

This is, of course, a placeholder term. "Executable essay" would in many ways be more natural, though unfortunately that term is perhaps even more naturally interpreted as meaning "web page", in the current context.

Tools for thought must be developed while doing serious work.

### The aspiration to canonical content

Seymour Papert, one of the principal creators of the Logo programming language, had a remarkable aspiration for Logo. Logo is sometimes described as a “programming language for children”, and people sometimes think Papert was mostly interested in helping children learn how to program. But that wasn’t Papert’s principal intent. Rather, Papert wanted to create an immersive environment – a kind of “Mathland” – in which children could be immersed in mathematical ideas. In essence, children could learn differential geometry by going to Mathland.

It’s a beautiful aspiration, and Logo contains many striking ideas. But as far as we know, no professional differential geometer (or, more generally, mathematician) uses Logo seriously as a tool in their work. And upon reflection that seems troubling. If Logo genuinely expresses the ideas of differential geometry, why don’t differential geometers use it? You start to wonder: might it be that Logo leaves out important ideas about differential geometry, maybe even the most important ideas about differential geometry? After all, while mathematically trained, Papert wasn’t himself an accomplished differential geometer. How would he even have known what to include? And certainly most of the people interested in Logo aren’t qualified to make that judgment.

There’s a standard retort to this, which we’ve heard from within the Logo community. It’s to talk about the “floor” and “ceiling” of different environments for thought. In this account, Logo has a low floor (meaning anyone can use it) and a low ceiling (so it’s not well adapted for the sort of advanced work a professional would want to do).

At first this sounds plausible. But upon reflection it’s difficult to make much sense of. How do the creators of Logo know that mastering Logo helps later with understanding real (forgive us!) differential geometry? What’s the criterion for success? One of us (MN) worked for several years doing research in the closely related field of Riemannian geometry. While Logo is enjoyable to use and contains many fun ideas, MN has trouble seeing that learning Logo would help much in learning differential geometry.

At the end of Norvig’s economics essay is a short afterword explaining how he came to write the essay. Shortly before writing the essay he’d heard about the kinds of economic models discussed in the notebook, and he wanted to explore several questions about them. After talking it over with some colleagues they decided to each independently attack the problems, and to compare notes.

Although Norvig’s essay is, in some sense, “educational”, Norvig’s intent was to explore a set of problems he himself was genuinely curious about. The educational aspect was a byproduct.

And so what you have is a world-class research scientist who wanted to explore a set of questions. He used the Jupyter medium to do those explorations, and then to share that exploration with the world. And he shared it in a form where others could immediately build upon and extend his thinking.

There’s a lot of work on tools for thought that takes the form of toys, or “educational” environments. Tools for writing that aren’t used by actual writers. Tools for mathematics that aren’t used by actual mathematicians. And so on. Even though the creators of such tools have good intentions, it’s difficult not to be suspicious of this pattern. It’s very easy to slip into a cargo cult mode, doing work that seems (say) mathematical, but which actually avoids engagement with the heart of the subject. Often the creators of these toys have not ever done serious original work in the subjects for which they are supposedly building tools. How can they know what needs to be included?

Concretely: suppose you want to build tools for subject X (say X = differential geometry). Unless you are deeply involved in practicing that subject, it’s going to be extremely difficult to build good tools. It’ll be much like trying to build new tools for carpentry without actually doing any carpentry yourself. This is perhaps part of why tools like *Mathematica* work quite well – the principal designer, Stephen Wolfram, has genuine research interests in mathematics and physics. Of course, not all parts of *Mathematica* work equally well; some parts feel like toys, and it seems likely those are the ones *not* being used seriously internal to the company.

There’s a general principle here: *good tools for thought arise mostly as a byproduct of doing original work on serious problems*. They tend either be created by the people doing that work, or by people working very closely to them, people who are genuinely bought in. Furthermore, the problems themselves are typically of intense personal interest to the problem-solvers. They’re not working on the problem for a paycheck; they’re working on it because they desperately want to know the answer.

A related argument has been made in Eric von Hippel’s book “Democratizing Innovation” (2005), which identifies many instances where what appears to be commercial product development is based in large or considerable part on innovations from users.

Many people have asked why we wrote our first mnemonic essay about quantum computing. If we'd chosen an easier subject we could have attracted a much larger audience. But we also wanted the essay to be authentic, to be about problems we wanted to solve. One of us (MN) has done a lot of original research work on quantum computing. The essay reflects that thinking. Indeed, the framing of the essay is about answering a question MN personally wanted to answer: if humans ever discovered aliens, would they have computers, and if so, what types of computer would they have? This perhaps sounds like a contrived question, but it's quite serious, and turns out to be a deep question with a nontrivial answer: writing the essay helped MN substantially improve his understanding of the question.

That said, answering that question wasn't the principal point of creating the essay: making the mnemonic medium was. And for future work on tools for thought, it'd be valuable to push much harder on questions we'd genuinely like to answer ourselves. That's a way of keeping yourself honest, ensuring you're not just building a flashy toy, but something genuinely useful for solving real problems that are of independent interest.

In serious mediums, there's a notion of *canonical media*. By this, we mean instances of the medium that expand its range, and set a new standard widely known amongst creators in that medium. For instance, *Citizen Kane*, *The Godfather*, and *2001* all expanded the range of film, and inspired later film makers. It's also true in new media. YouTubers like Grant Sanderson have created canonical videos: they expand the range of what people think is possible in the video form. And something like the *Feynman Lectures on Physics* does it for textbooks. In each case one gets the sense of people deeply committed to what they're doing. In many of his lectures it's obvious that Feynman isn't just educating: he's reporting the results of a lifelong personal obsession with understanding how the world works. It's thrilling, and it expands the form.

We've been disappointed by how unambitious people are in this sense with *Jupyter* notebooks. They haven't pushed the medium all that hard; there is no *Citizen Kane* of Jupyter notebooks. Indeed, we're barely beyond the Lumière brothers. Examples like Norvig's notebook are fine work, but seem disappointing when evaluated as leading examples of the medium.

Aspiring to canonicity, one fun project would be to take the most recent IPCC climate assessment report (perhaps starting with a small part), and develop a version which is executable. Instead of a report full of assertions and

references, you'd have a live climate model – actually, many interrelated models – for people to explore. If it was good enough, people would teach classes from it; if it was really superb, not only would they teach classes from it, it could perhaps become the creative working environment for many climate scientists.

One promising exploration in this direction is *The Structure and Interpretation of Classical Mechanics*, a beautiful executable book building up classical mechanics. Many theorems of classical mechanics aren't just expressed in static form, on the page, but live, as code which can be modified by the user. Theorems become APIs, which can literally be applied to other objects, and chained together. It uses a much more powerful underlying model than Jupyter, developing a new symbolic language as part of the book. It has many flaws – among them, the book doesn't run live in the browser, making it difficult for users to experiment. And while the book is well written, the authors do not understand classical mechanics as deeply as the authors of some other books. But it's nonetheless an inspiring evocation of what is possible. And it hints at what is possible when authors use executable books for a serious purpose, and aspire toward canonical media.

### Stronger emotional connection through an inverted writing structure

Consider an author writing a popular book about quantum mechanics. Such an author is in a strong position: they can begin their book with astonishing phenomena such as black hole evaporation, quantum teleportation, and the role of quantum fluctuations in the early universe. Or, if they wish, they can start with some of the deepest mysteries known to humanity: the relationship between quantum mechanics and gravity, or the quantum measurement problem. There is no shortage of extraordinary phenomena and beautiful mysteries. These are the kind of things which touch a chord inside many, perhaps most people. And so it's relatively easy to draw readers in, to get them engaged, and keep them connected.

By contrast, consider a typical technical book about quantum mechanics. It's very unlikely to start with black hole evaporation or quantum teleportation – and if it does, such a discussion will be perfunctory. Instead, it will start out drily, with technical minutiae. Complex numbers. Wavefunctions. Many different types of differential equations, and how to solve them. Hermitian and unitary operators. And so on, piece by piece slowly building up all the machinery needed to solve quantum mechanical problems. It may be tens or even hundreds of pages before

the book begins to connect to the exciting problems which form the bread-and-butter of popular accounts.

People with little experience doing good technical writing often complain about this dry, bottom-up approach. They will complain that writers should instead stay closer to the fun material, and use less technical notation and nomenclature. But when competent writers attempt to follow this prescription, invariably it works poorly.

One problem is that a person can spend years reading analogies about black hole evaporation, quantum teleportation, and so on. And at the end of all that reading they typically have... not much genuine understanding to show for it. The analogies and heuristic reasoning simply don't go far. They may be entertaining and produce some feeling of understanding. But the reasoning won't scale out; it can't be applied to other phenomena, at least not without lots of caveats, caveats the reader is in no position to understand or apply. As a result, good technical writers instead mostly build things up from first principles, with occasional digressions to the broader motivating picture. And that means starting with a lot of detailed, technical minutiae.

It's striking to contrast conventional technical books with the possibilities enabled by executable books. You can imagine starting an executable book with, say, quantum teleportation, right on the first page. You'd provide an interface – perhaps a library is imported – that would let users teleport quantum systems immediately. They could experiment with different parts of the quantum teleportation protocol, illustrating immediately the most striking ideas about it. The user wouldn't necessarily understand all that was going on. But they'd begin to internalize an accurate picture of the meaning of teleportation. And over time, at leisure, the author could unpack some of what might *a priori* seem to be the drier details. Except by that point the reader will be bought into those details, and they won't be so dry.

In other words, you could begin an executable book with material the users already care about, can connect to easily, and find motivating. For instance, you could begin by exploring teleportation or the Big Bang. But such an opening won't suffer the drawback of popular science, of being vague and imprecise. Rather, the interface would be completely well specified. And, with some care, the interface could be scaled out, applied in ever-expanding contexts. The understanding would be transferable. Even a

user who has understood only a tiny part of the material could begin tinkering, building up an understanding based on play and exploration. It's common to dismiss such an approach as leading to a toy understanding; we believe, on the contrary, that with well enough designed scaffolding it can lead to a deep understanding. Developed in enough depth, such an environment may even be used to explore novel research ideas. To our knowledge this kind of project has never been seriously pursued. But it'd be fun to try.

The masters of this are video game designers. See, for example, Dan Cook, [Building a Princess Saving App](#) (2008), and Jonathan Blow and Marc ten Bosch, [Designing to Reveal the Nature of the Universe](#) (2011).

## Summary and Conclusion

We've covered a lot, and it's helpful to distill the main takeaways – general principles, questions, beliefs, and aspirations. Let's begin with memory systems, particularly the mnemonic medium:

- **Memory systems make memory into a choice, rather than an event left up to chance:** This changes the relationship to what we're learning, reduces worry, and frees up attention to focus on other kinds of learning, including conceptual, problem-solving, and creative.
- **Memory systems are in their infancy:** it is possible to increase effective human memory by an order of magnitude, even beyond what existing memory systems do; and systems such as the mnemonic medium may help expand the range of subjects users can comprehend at all.
- **What would a virtuoso use of the mnemonic medium look like?** There's some sense in which the mnemonic medium is "just" flash cards. The right conclusion isn't that it's therefore trivial; it's that flash cards are greatly underrated. In writing *Quantum Country* we treated the writing of the cards with reverence; ideally, authors would take card writing as seriously as Nabokov took sentence writing. Of course, we didn't reach that level, but the aspiration expands the reach of the medium. What would virtuoso or even canonical uses of the mnemonic medium look like?
- **Memory systems can be used to build genuine conceptual understanding, not just learn facts:** In *Quantum Country* we achieve this in part through the aspiration to virtuoso card writing, and in part through a narrative embedding of spaced repetition that gradually builds context and understanding.

- **Mnemonic techniques such as memory palaces are great, but not versatile enough to build genuine conceptual understanding:** Such techniques are very specialized, and emphasize artificial connections, not the inherent connections present in much conceptual knowledge. The mnemonic techniques are, however, useful for bootstrapping knowledge with an *ad hoc* structure.
  - **Memory is far more important than people tend to think:** It plays a role in nearly every part of cognition, including problem-solving, creative work, and meta-cognition. The flip side is that memory systems themselves want to grow into other types of tools – tools for reading, tools for problem-solving, tools for creating, tools for attention management. That said, we don't yet know what memory systems want to be. To reiterate: memory systems are in their infancy.
- The mnemonic medium is merely one prototype tool for thought. We also discussed several other ideas, including mnemonic video and executable books. Here are some key takeaways:
- **What practices would lead to tools for thought as transformative as Hindu-Arabic numerals? And in what ways does modern design practice and tech industry product practice fall short?** To be successful, you need an *insight-through-making loop* to be operating at full throttle, combining the best of deep research culture with the best of Silicon Valley product culture.
  - **Tools for thought are (mostly) public goods, and as a result are undersupplied:** That said, there are closely-related models of production which have succeeded (the games industry, Adobe, AutoDesk, Pixar). These models should be studied, emulated where possible, and used as inspiration to find more such models.
  - **Take emotion seriously:** Historically, work on tools for thought has focused principally on cognition; much of the work has been stuck in Spock-space. But it should take emotion as seriously as the best musicians, movie directors, and video game designers. Mnemonic video is a promising vehicle for such explorations, possibly combining both deep emotional connection with the detailed intellectual mastery the mnemonic medium aspires toward.
  - **Tools for thought must be developed in tandem with deep, original creative work:** Much work on tools for thought focuses on toy problems and toy environments.

This is useful when prototyping, but to be successful such tools must ultimately be used to do serious, original creative work. That's a baseline litmus test for whether the tools are genuinely working, or merely telling a good story. Ideally, for any such tool there will be a stream of canonical media expanding the form, and entering the consciousness of other creators.

Let's return to the question that began the essay: how to build transformative tools for thought? Of course, we haven't even precisely defined what such transformative tools are! But they're the kind of tools where relatively low cost changes in practice produce transformative changes in outcome – non-linear returns and qualitative shifts in thinking. This is in contrast with the usual situation, where a small change in practice causes a small change in results.

Historically, humans have invented many such transformative tools for thought. Writing and music are ancient examples; in modern times, tools such as *Photoshop* and *AutoCAD* qualify. Although it's very early days, we believe the mnemonic medium shows much promise. It needs to be developed much further, along the lines we've described, and likely requires additional powerful ideas. But we believe it's possible for humanity to have a widespread memory practice that radically changes the way we think.

More broadly, we hope the principles in this essay will help support the creation of more transformative tools for thought. Historically, most invention of tools for thought has been done bespoke, by inspired individuals and groups. But we believe that in the future there will be an established community that routinely does this kind of invention.

## Acknowledgments

This essay is based on conversations with many people. Particular thanks to David Albert, Shan Carter, May-Li Khoe, Robert Ochshorn, Grant Sanderson, Caitlin Sikora, and Bret Victor. Thanks also to Nicky Case for feedback on a draft of the essay. This work was supported by our Patreons. AM is supported in part by a grant from Emergent Ventures. MN's work was supported by YC Research.

## Citation

For attribution in academic contexts, please cite this work as:

Andy Matuschak and Michael Nielsen, "How can we develop transformative tools for thought?",  
<https://numinous.productions/tft>, San Francisco (2019).

Authors are listed alphabetically.

## License

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).