🎉 Andy Matuschak just shared "Ratcheting progress in tools for thought" for patrons only

Andy Matuschak

Ratcheting progress in tools for thought

There are some people trying to develop tools for thought, but there isn't yet a meaningful *field* around tools for thought. The difference is that a field is about ratcheting: developing a growing shared corpus of general knowledge and methods which allow projects to meaningfully build on each other, across researchers and across years, on and on in an upward cycle. Individuals here and there have contributed powerful insights, certainly, but true success for this field would mean a new practice of human tool-making, and the creation of many tools which transform what people can think and do.

Michael Nielsen and I have suggested that the most powerful tools for thought express deep insights into the underlying subject matter. Creating them involves what we call "insight through making," in which powerful subject-matter ideas enable new systems, and observations of those systems enable new insights, and so on. An idealized cycle of activity might involve something like these key steps:

1. **Identifying powerful insights** about some subject domain or about cognition in general which might be fruitfully systematized
2. **Building systems** which express those insights in their primitives
3. **Observing serious use** of those systems in authentic contexts, and of your theoretical insights refracted through them
4. **Distilling generalizable insight** from those observations which produce new understanding about the subject domain or cognition, and which permit new, better systems to be built
5. **Disseminating** that insight so that others can build on it

This high-level view glosses over a great many details, and every one of these steps is a complex practice in which one can build skill over decades. But we can see these practices in the "golden era" design work of computer-powered tools for thought: Engelbart's NLS, PARC's Alto, Sutherland's Sketchpad, and

so on. Unfortunately, if we look at the contemporary proto-field, we'll find that most people interested in tools for thought (myself included) are not reliably performing all these steps—which has left our struggling field without a functioning ratchet.

Common failure modes

One failure mode, particularly common in academia, comes from lacking a serious context of use. Often this just means that the observation step centers on misleading artificial environments. Sometimes this happens because researchers don't invest seriously enough in the system being built, which makes it difficult to distill powerful new insights from what they've made. But more subtly, without a serious context of use driving the research project, their initial insights are likely to be limited or misdirected.

Startups and tech businesses are powerful venues for tool development. They're generally not *trying* to push the field of tools for thought forward. But we might hope that they happen to push it forward anyway. Unfortunately, a few common patterns prevent most tech industry efforts from contributing to a ratcheting field.

Perhaps the most common pattern of all is that people in the tech industry focus mostly on building systems. Those systems are usually expressions of technological or market insights, rather than of fundamental insights about a subject domain or about cognition. That's not a problem as far as the business or its users is concerned! But if their systems don't reify new ideas, we can't draw much field-level insight from observation.

Another common pattern for tech companies is that they're *founded* on the premise of some powerful insights, insights which motivated the founders to start a business (or for an existing business to conceive a new product line). The company instantiates those insights in systems, observe them in serious contexts, distill new ideas from observation, and improve the product. This is great! But the object of all this iteration is rarely "generalizable subject-matter insight," and for good reason. These businesses are trying to improve their product's performance in its customers and market. Sometimes we get lucky, and this happens to produce generalizable insights—typically only after others reverse-engineer and disseminate them. But often this difference in focus means a narrower scope, as far as the field is concerned. They're fixing pain points, adding line extensions, smoothing workflows, adding features… generally without changing any of the foundational theory of the product. In fact, changes to the foundational theory of the product are usually (and often rightly) "off limits", or just not even salient for product teams. I think good tools-for-thought research often focuses on transcending and discarding the current system, asking "how should we build the next system?". But good business

usually don't throw out their core product and build a meaningfully different one every few years.

Dissemination may be another challenge for tech companies contributing to this field. It's fascinating to see companies like AutoDesk, Adobe, and Epic Games publish incredible papers about fundamental problems in computer graphics. But if tools-for-thought-ish companies are producing powerful new theories about cognition or their subject matter, we rarely see those ideas published.

Some of my favorite work in tools for thought comes from idiosyncratic Twitter tinkerers. This group often produces fascinating work, but it's usually missing one or more of these steps. The most common pattern seems to be: a bricoleur identify some powerful idea about a representation and design a prototype, but then fails to engage seriously with observing and deriving insight from the systems they've built. Sometimes this comes from technical barriers—the prototype is too quick-and-dirty to be used in a serious context, so their insight is limited. But I think there's also a cultural gap here, a missing research practice of careful, diligent observation and synthesis. Too often these projects have the flavor of "Look, I made a thing! Isn't it cool? How many people can I get to use it?". But the question they need to be answering is: "What powerful, generalizable ideas can we learn from this project? How should the next wave of systems build on this?"

Designing research insight production into the system

Institutional incentives aside, one reason it's so rare to find this cycle operating smoothly is that it's incredibly difficult to do. The steps are interdependent. You can't just have a powerful idea, then design a system which expresses it, then observe it, and so on. You have to design the system and manipulate it in a way which will reveal insight about the idea the system represent. Or, to put it another way, the system has to be shaped in a way which allows you to ask the questions you want to ask. But often you can't even identify the right questions to ask before you see the system in operation!

For instance, one of the key ideas behind Quantum Country is that authors can help readers deeply internalize complex, abstract topics by interleaving narrative and retrieval practice. We built a system which expresses that idea. We have lots of data. Readers seem to feel it works. We can see that people do indeed retain the material they read. But what should the field learn from this experiment? What generalizable conclusions can we draw? How can we improve our understanding of that initial idea so that we can create a better future medium?

Sometimes interesting answers come by accident. For instance, when interviewing readers about their experiences, we were surprised to discover that memory effects aside, the regular review sessions meaningfully changed how people related to the material. It caused them to think of themselves as "doing

quantum computing" in a much more serious way than they would if they'd just read some essay on one afternoon. That insight gave rise to Timeful Texts and other related directions.

But many insights can't be explored through passive observation and open-ended interviews. How exactly do the narrative and the retrieval practice interact? Is the main effect convenience—i.e. the prompts are delivered while reading, whereas if they came later, you wouldn't bother? Or is it meta-cognitive —i.e. the prompts' feedback regulate your reading, causing you to re-read passages where you fared poorly? Or is it mostly about memory—i.e. the most efficient retrieval practice schedule involves practicing and reinforcing knowledge immediately, rather than a few days later? Or something else entirely? Different answers to these questions would point to substantially different paths for the evolution of the mnemonic medium.

To answer these questions, the system has to be designed in a way which produces the necessary observations. Or you have to manipulate the system with an experiment, which may be difficult if you didn't initially architect your system with those questions in mind. Somewhat more subtly, I've found that designing tools-for-thought experiments such that the results might tell you something possibly *generalizable* is particularly difficult—a contorted balancing act of theory, interface design, engineering, and experimental methods. It's like what cognitive psychologists have to deal with in their experimental design, except the "apparatus" is a system which must both solve real-world problems and *also* produce the necessary experimental data.

Ben Shneiderman, a pioneering human-computer interaction researcher, offers this charming schematic for research project design in *The New ABCs of Research*. He calls it the "two parents, three children" pattern.◆
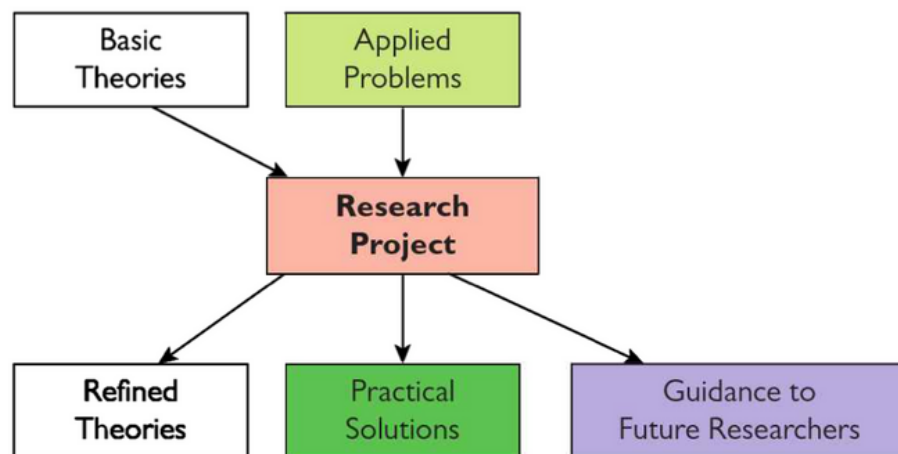

Figure 1.5 Two parents, three children: A pattern for organizing a research project.

The challenge is similar to what learning scientists must do in designing educational interventions. In *Principles and Methods of Development Research*,

Jan van den Akker offers a beautiful distillation of what a unit of progress looks like in that field (thanks to Sarah Lim for the pointer):

> [Educational design] principles are usually heuristic statements of a format such as: "If you want to design intervention X (for the purpose/function Y in context Z), then you are best advised to give that intervention the characteristics A, B, and C (substantive emphasis), and to do that via procedures K, L, and M (procedural emphasis), because of arguments P, Q, and R [(theoretical emphasis)]."

The key thing it does is to explicitly connect the dots between a grounded theoretical claim, the implied design approach, and the desired outcome. I'm certainly wary of trying to fit all research into some kind of formula like this, but how clarifying it is to have this target painted so sharply! If you're a researcher and you want to develop some new intervention, you need to design an experiment whose results can generate a statement of this kind.

I think that research cycles in tools for thought should strive to generate analogous statements. What are the consequences of our theories and our design decisions, the consequences which others can build on? Progress will mean being able to make lots and lots of fine-grained heuristic statements like "Retrieval practice systems should offer users the opportunity to retry a few minutes later when they fail to remember an answer [because I've run a controlled experiment and found that without that opportunity, lapses are about 10pp more likely to persist on the subsequent attempt]." But it will also important to be able to make big-picture statements about core mechanisms of systems, like "Intermittent follow-on review sessions can be used to change readers' emotional relationship to written material in ways A, B, and C through authorial methods X, Y, and Z, as predicted by theory P, Q, and R and supported by experiments K, L, and M."

Or, to take another example, I know many of my readers are fans of outline-based text editing. This morning, inspired by a message from patron Ethan Plante, I went looking for academic work on the theoretical or empirical foundations of outline processors. I was shocked how little I could find. So, if you're experimenting with building outline processors, or "block-based" tools, or whatever, some questions to be answered: what effects do these alternative writing primitives have on composition? on thinking? on reading? What is the theory which would explain these effects? How would we know if it were true? What else does that theory imply? What would research systems which could answer these questions look like, and how are those different from the commercial systems people build?

Some positive examples

I wrote this post to clarify my own challenges, so it's necessarily coming from a place of frustration. I want to close on a more positive note, by pointing to a few contemporary examples of projects which complete the full cycle I've described:

Bret Victor's projects are the classic modern example. I'll give a too-short summary of one branch of research. Explorable Explanations (2011) proposes that a reading environment could become *an environment to think in* if authors didn't just present data, but rather embedded their live models into the written medium. A sequence of projects on reading, writing, and interacting with dynamic systems followed, pursuing this line of thinking and several adjacent ones, each contributing substantive generalized insight. Eventually Bret and colleagues became dissatisfied with the limitations of screens and the asymmetric role of authors, leading to the wildly original concepts behind his current project, Dynamicland—a building that is a computer.

Ink and Switch, the industrial research lab, did a thoughtful and well-documented series of experiments with freeform multimodal tablet interfaces for supporting creative thought. Their first, Capstone, was based on a model of creative work as dependent on gathering and sifting raw material for patterns and insights. They built a number of interactions to support their model, identified opportunities and limitations with that system, and designed a new system based on those insights called Muse. That project put inking front-and-center and produced general ideas about designing ink-centric interfaces with no chrome. The research project is now a product company (co-founded by patron Adam Wiggins). I'm excited for their attempt to prove out a translational R&D model, and I'll be interested to see whether they're still able to generate and disseminate generalizable insights in that context.

Piotr Wozniak, the contemporary founder of spaced repetition, is another great example. He had the original key insight that a computerized system for spaced repetition could lead to very large, very cheap memory databases. But he's been iterating on those ideas for decades now. He's not just been optimizing the scheduling algorithm but using the data as part of proposing and exploring new models of human memory (e.g. How much knowledge can human brain hold). This research doesn't seem to have been constrained by SuperMemo's commercial setting, though that's perhaps because it's not run with the intensity of a modern U.S. software company.

Evan Wallace at Figma developed and documented a new primitive for representing and editing vector paths, motivated by practical problems with existing vector pen tools, which more directly (naively?) expose the underlying Bezier curves. He shows how this new representation makes certain common operations easier to perform. This work may not transformatively change and expand the thoughts people can think, but both Sketchpad and Illustrator's original pen tool were certainly significant, and this seems to meaningfully extend that work. I do wish Evan had written about the work more substantively,

but of course, he's trying to improve a product, not contribute to a field. There's a nice recent technical write-up, but it's implementation-focused.

In the sphere of para-academic Twitter tinkerers, I want to applaud Omar Rizwan's experiments with TabFS. That project expresses a deep insight of Omar's: that a shortcut to end-user programming may lie in extending the architecture of operating systems up to application-level objects—like browser tabs. In many ways, this project is an extension of Plan 9, but with a powerful injection of worse-is-better folk/craft philosophy. But unlike many of my beloved Twitter tinkerers, Omar's been diligently synthesizing and disseminating new insights produced by how he and others have been using TabFS. Unfortunately, those insights are in sponsor email newsletters which lack permalinks, but you can get a sense from his Twitter. Longer-term, I'm sure Omar will produce some durable write-up of what he learns, something which others can build on—he's done it for past work.

What are your favorite contemporary examples of people who are completing "full cycles" of work in tools for thought? Please share them in the comments.

View on Patreon

Get the Patreon app

Download on the App Store   GET IT ON Google Play