

From the impossible to the improbable

A Private ARPA user guide



How can we enable more science fiction to become reality?

Looking to outliers who have done it before is a good place to start. [After digging into why DARPA works](#), I asked the natural follow-up question: how could you follow DARPA's narrow path in a world very different from the one that created it?

This piece is my answer. A voluntary hybrid for-/nonprofit organization that leverages empowered program managers and externalized research to shepherd technology that is too researchy for a startup and too engineering-heavy for academia. Then, I lay out a precise roadmap for how to make it happen.

However, in this case a naked plan is not enough. It violently violates the

historically-validated principle to change only one thing at a time; the burden of proof rests heavily on me to explain why more incremental approaches just won't cut it. The proposal doesn't stand on its own -- it needs a foundation of evidence and argument. This foundation spans many topics: from the role of value capture in technology creation, to the now-defunct historical role of industrial labs, to tactics for institutional longevity, and the nitty-gritties of how to fund operations and why most options suck.

No single organization can enable science fiction to become reality alone. Given that fact, I don't want to only propose a single organization. More ambitiously, I want to provide a user manual for others to build DARPA-riffs and perhaps innovation organizations we haven't yet dreamed of.

Take home messages:

1. The niche in the innovation ecosystem once occupied by industrial labs is unfilled — academia, startups, and modern corporate R&D do not cut it.
2. It is possible to build a private organization that riffs on DARPA's model to fill this niche.
3. The nature of the model, value capture, and the nature of the work itself means that this organization will not be profitable as a whole, but it may be possible for it to leverage for-profit investment.
4. Private DARPA (PARPA) needs to play the long game; it will derisk a series of hypotheses and go through three major evolutionary phases before it looks like its namesake.
5. There are many fundamental tensions and incentive traps surrounding the path to building any innovation organization. Laying them out as explicitly and precisely as possible may enable PARPA and other organizations to sail past them safely.

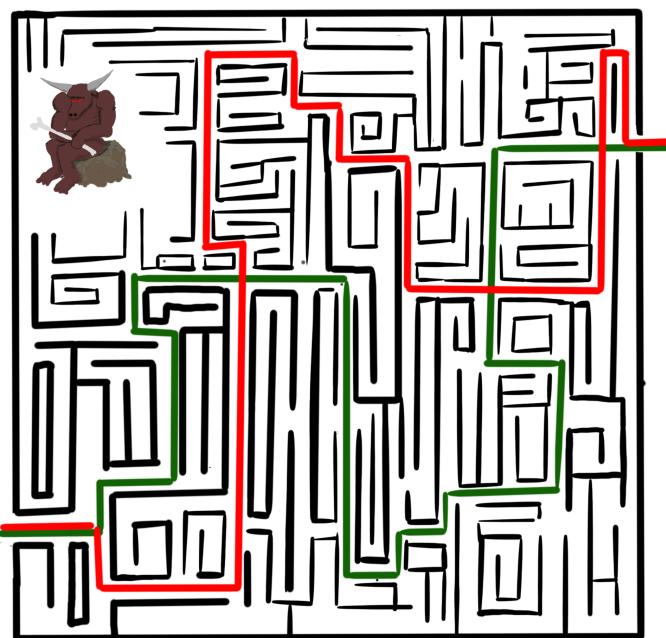
Distillation

This section compresses the entire piece into a short essay. Each sentence roughly corresponds to a section, though I haven't linked all of them directly. I recommend reading through, noting where you find yourself disagreeing with something or wanting more detail and then reading the corresponding full section.

Any new organization needs to navigate an 'idea maze' made of many often-irreversible decisions; it's full of dead-ends and existential threats (like minotaurs!). The less precedent there is for an organization, the more daunting the idea maze; fewer heroes who traversed similar mazes can describe the forking paths you will face. There is precious little precedent for riffing on DARPA model! So in addition to sketching my plan to navigate the maze, it's important to establish why the maze exists in the first place and its general layout to both frame the specific plan and enable other people to make their own.

PARPA - how I propose to navigate the maze

PARPA



P(private)ARPA will tackle work that's too researchy for a startup and too engineering heavy for academia. It will go after this 'Solutions R&D' niche by

taking the following high-level steps:

1. Create and stress-test unintuitive research programs in a systematic (and therefore repeatable) way.
2. Use that creditability to run a handful of research programs and produce results that wouldn't happen otherwise.
3. Use that creditability to run more research programs and help them 'graduate' to effective next steps.
4. Make the entire cycle eventually-self-catalyzing by plowing windfalls into an endowment.

Each step comprises a bundle of hypotheses — “I suspect we should go right and walk twenty yards” in idea-maze terms. These hypotheses are both sequential and hierarchical, eventually getting down to near-implicit ones like “if I send an email to professor so-and-so, we will have a meeting and they will be able to point me to two other people to talk to.” I’ll leave that level of detail to your imagination but the first-and-second level hypotheses are:

1. It’s possible to design programs in a systematic (and therefore repeatable) way.
 - a. It’s possible to find people who want to do pieces of work that would not happen otherwise.
 - b. There exist several areas that could possibly yield results (defined loosely) in 3-5 years (given steady work).
 - c. For each potential program, it’s possible to come up with a set of small experiments that could further confirm or deny its feasibility within 12-24 months and ~\$1M scale.
 - d. It’s possible to get people to undertake seedling experiments.
 - e. Accomplishing 1a->1d will involve a combination of 1-1 conversations driven by reading papers, finding gaps, reaching out

- to people, and holding small, intensive workshops.
- f. It's possible to find people besides myself who are excited to be PMs and actually *do* 1a->1e.
 - g. It's possible to systematize this process.
2. It's possible to execute on these programs in a way that relaxes academic / startup constraints.
 - a. It will be possible to coordinate several research projects towards a coherent goal.
 - b. The early pieces of work will be done by some combination of academic labs, contract R&D orgs, independent researchers, and possibly people within corporate R&D who are able to do work for grants.
 - c. People will be willing to shift their organizational affiliation to a temporary organization devoted to the program as it progresses.
 3. It's possible to 'graduate' programs in a way that gives them a life of their own.
 - a. During the programs themselves it will be possible to do some of the pre-work to figure out the best way to graduate technology in order to maximize its beneficial impact on the world.
 - b. At least some of the people who have been working on the program will want to carry it forward in some form.
 - c. Some of the technology will make sense as a company, some of it will make sense as a non profit, some of it will make sense as a licensed technology, and some of it will make sense as open source.
 4. It's possible for this entire cycle to eventually become self-catalyzing
 - a. An early (non-monetary) component of self-catalyzation is a

- community that generates good inbound programs.
- b. “Pseudo-self-catalyzation” is state where the organization is getting enough consistent revenue through some combination of donations, spinoffs, and other sources like licensing and consulting that it can continue to run multiple programs assuming those revenue sources continue.
 - c. Full self-catalyzation can only happen by building up an endowment.

What’s the best organizational structure to support these hypotheses? It’s often an afterthought, but organizational structure is critical to institutional capacity. A (big N) Nonprofit that owns a chunk of a for-profit will allow PARPA to explicitly acknowledge that it will be creating hard-to-capture value but it’s likely that it will create capturable value as well; creating companies is *sometimes* the most effective way to get technology out into the world.

It’s important for big things to start small and PARPA is no exception. While it’s growing, certain organizational sizes and scopes are more ‘stable’ than others; with a \$50M/year budget PARPA won’t simply do 10x more of the same things that it would do with \$5M/year, and \$25M/year would be insufficient for the former scope and excessive for the latter. In light of this, PARPA will go through three evolutionary ‘stages’ gated both by available funds and whether we feel like the previous stage has adequately derisked the shift.

Problems and Possibilities — the boundaries of the maze and the dangers within

Problems and Possibilities



We've established what PARPA is and how it's going to tackle solutions R&D. That structure and those hypotheses aren't arbitrary — the path is based on a very specific map of the idea maze. Incentives and constraints *matter*. Drawing the map in addition to sketching the path is important for two reasons. First, without the map, choices can seem fairly arbitrary and relatively fungible; instead, they're weaving a narrow path between many incentive traps and fundamental tensions. Second, there isn't just one path through the maze; the map will ideally enable others to chart their own path and build different flavors of DARPA-riffs or other organizations entirely.

Before we fill in the twists and turns of the maze, it's important to ask "where are the its boundaries and why does it exist in the first place?" Perhaps the best way to reason about mazes is to interrogate those who have made similar journeys; teasing out what made past organizations successful and why current institutions seem to fall flat.

Golden age industrial labs like Bell Labs, GE Labs, Kodak Research, Pfizer Research, and more stand out as some of the most successful solutions R&D

organizations. They played a major role in technologies at the core of the modern world — from transistors and plastics to lasers and penicillin.

What made golden age industrial labs good at solutions R&D? Obviously this question is impossible to answer definitively but we can identify some telling patterns. They enabled work on general-purpose technology for a long time before it needed to specialize. They enabled playful research that at the same time was oriented towards useful outcomes. Golden age labs provided an environment that enabled projects to ramp smoothly from eye-twinkles to mass-manufactured products; having a ‘default customer’ helped the smooth ramp, as did bringing together different disciplines — especially having manufacturing experts present early in the prototyping process.

Studying industrial labs also leads to a sobering conclusion: The constraints on corporate R&D today make building a new organization modeled directly on Bell Labs a fool’s errand. Any innovation organization needs an aligned ‘money factory’ to provide the operating capital to support high-uncertainty projects with long time horizons. Through this lens, it’s notable that golden age industrial labs satisfied a trifecta of conditions that, outside of a few exceptions like AI research,¹ are no longer the case:

1. They were run by a monopoly.
1. They were working on a clearly high potential technology.
2. That technology addressed one or more existential threats to the company.²

¹ Notably, AI technology still fulfills all these conditions and conveniently for our narrative, AI labs like Google Brain, DeepMind, and FAIR seem to be the exceptions to the decline of industrial labs!

² Condition #3 was originally “1+2 are insufficient if the technology the lab works on isn’t tied into the company’s core business.” However, this isn’t quite right. There are situations where technology work on the core product can have nothing to do with existential risks to the business (for example Salesforce.) Similarly there are situations where technology work has little to do with the core product but addresses an existential

It's reasonable to ask whether other existing institutional structures in the innovation ecosystem can, perhaps with some tweaks, fill the solutions R&D niche. Academia resembles industrial labs in many ways. Indeed, cultural expectations to produce new technology have shifted from industrial labs to academia over the past several decades. However, academia incentivizes scientific inquiry over engineering design. As a result, it can't muster the level of focused engineering work needed to push most technology beyond nifty demos.

Once a technology requires more focused engineering work than an academic lab can support it's common to spin it out into a startup company. Startups have their own constraints, however. The central factor driving these constraints is that startups need to grow fast and be (eventually) profit-focused. The need to grow quickly is at odds with the longer-term research that new technologies often need.

Profit exerts a gravitational pull on any conversation about innovation. It's important for both practical and ideological reasons; even ignoring wealth, profit makes organizations auto-catalyzing which in turn lengthens organizational longevity and (assuming competence!) an organization's chances of doing great things. On top of the fact that it is just plain hard to capture value from research, there is a significant class of innovations that would have created drastically less value for the world if their creators had tried to capture that value. Profit-maximizing organizational structures can hamstring technology development.

Unfortunately, Nonprofits have their own problems. Fundraising requirements and donor incentives mean that the traditional nonprofit has problems supporting work that requires a significant chunk of capital and produces uncertain results on long timescales.

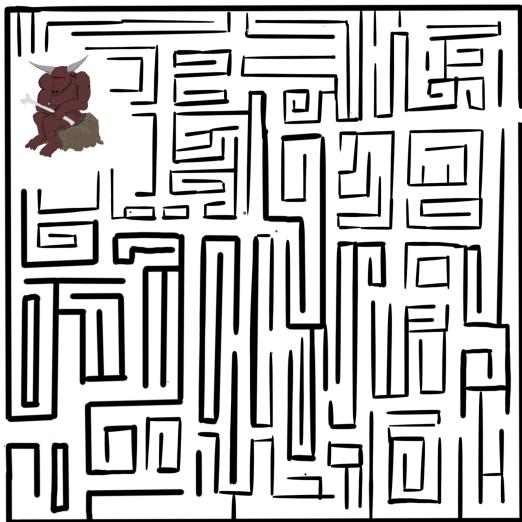
Together, the constraints on modern corporate research, academia, and startups create an institutional gap around things that are not paper-worthy enough for

threat (for example flashy research at Bell Labs keeping regulators from breaking up the monopoly.)

academia and not product-focused enough for a startup. The edges of this gap define the boundaries of the idea maze PARPA must navigate.

It is worthwhile and possible to riff on DARPA - the layout of the maze

Riffing on DARPA



Accelerating technological progress will require new institutional structures. Starting from a blank slate is a daunting task, that I propose we avoid. In addition to the golden age industrial labs, DARPA stands out as another paragon of solutions R&D.³ Unlike Bell Labs though, it seems possible to riff on DARPA's model in the 21st century.

Despite programs in everything from material science to short-turnaround antibody generation and automated scientific literature synthesis, there are still plenty of potential programs that DARPA doesn't pursue. Many ideas are not sufficiently national defense-related and (shockingly!) DARPA isn't all-seeing and all-knowing. While an organization patterned off of Bell Labs can't thrive separate from a monopoly, there is no reason that an organization riffing on DARPA's model can't thrive separate from a government. And in fact decoupling from governmental constraints may be a boon.

There are a set of junctures that any institution riffing on the DARPA Model must

³ For a deeper dive on DARPA and a refresher on what the DARPA model entails see Why DARPA Works

traverse. At some of these junctures there's a clear choice for every DARPA-riff and at others it's unclear whether to go left or right. Some junctures are pure speculation: they might be completely avoidable but at the same time, they might make the difference between success and failure. It's important to lay out these possibilities both so that you can make your own judgements and to shine light on the paths PARPA will not take.

Things any DARPA-riff will need to consider

Any organization not built around a core of powerful but short-tenured program managers can barely call itself a DARPA-riff. Program managers' short tenures means that you need to think carefully about both their incentives and how to foster a strong intergenerational culture. The nature of their role also means it's almost impossible to hire someone who has done it before. This distinct absence of direct experience means that it's important to think about which other characteristics suggest someone might be a good program manager.

The importance of program managers belies the other core piece of the DARPA model: the programs themselves. The concept of a 'research program' is admittedly nebulous — I liken it to putting together a lego set where you need to create both the instruction book and many missing pieces!

I suspect that any successful DARPA-riff will require a legal structure that diverges from your traditional cast of characters eg. a simple Delaware C-Corp, LLC, 501(c)(3), etc. Specifically, the goal of the legal structure should be to invert the normal relationship between an innovation organization and its money factory, with the innovation organization in the driver's seat. For this inversion to work, it's important that the legal structure of a DARPA-riff give the most power to the people with the longest time horizons.

"How does the money work?" is perhaps the most critical question any DARPA-riff needs to ask. At the end of the day, the point of the organization is its research output. However, doing that work is impossible without (a good chunk of) money. And it won't necessarily pay for itself. Focusing on shorter-term income, licensing IP is a common answer to questions of how to pay for

technology development. Unfortunately, outside of a few specific areas like pharmaceuticals, licensing is unlikely to be a viable primary income source. There are many other tactics, each with its own constraints and experimental possibilities: donations, contracting, consulting, spinning out companies, etc. In the long run, an endowment that funds operations out of the interest on a large principle is the ideal money factory for a DARPA-riff. The goal of an endowment is both worth striving for and at the same time it's the organizational equivalent of "well first I need to have F* you money before I work on the thing I really want to."

Speculation

Any DARPA-riff will need to run several organizational experiments: in a way part of its role is as a serious context-of-use for experiments in research management. Some threads to pull on include:

- There is quite a bit of room to experiment with how programs are structured and organized. It is worth seriously considering how to create a program design discipline that enables better research and development.
- Structurally, creating temporary organizations for each program could enable a DARPA-riff to bridge the gap between external and internalized work and provide flexibility around research outcomes.
- Experiments to maintain organizational focus over a long period of time are underway; two that seem especially worth pursing are employee ownership and literally having a constitution with checks and balances.

Conclusion

Most people who entered the labyrinth — even those with a plan — didn't survive. Uncomfortably, the line between success and failure⁴ is fuzzy, which makes it especially important to explicitly examine what constitutes both success and failure. Laying out nebulous indicators of success and failure not only helps

⁴ Theseus killed the minotaur but ultimately caused his father to commit suicide. Success? Failure?

steer towards the former and away from the latter but also enables us to be as rigorous about the process as possible. Called shots allow us to collectively learn real lessons regardless of outcome instead of being dominated by post-hoc narratives.

I hope I've convinced you that it's at least worthwhile to try to build PARPA and its mission to *shift technology from impossible to inevitable*.

Introduction

The world is changed... Much that once was is lost, for none now live who remember it.

— Galadriel

When tracked against the admittedly lofty hopes of the 1950s and 1960s, technological progress has fallen short in many domains. Consider the most literal instance of non-acceleration: We are no longer moving faster. The centuries-long acceleration of travel speeds — from ever-faster sailing ships in the 16th through 18th centuries, to the advent of ever-faster railroads in the 19th century, and ever-faster cars and airplanes in the 20th century — reversed with the decommissioning of the Concorde in 2003, to say nothing of the nightmarish delays caused by strikingly low-tech post-9/11 airport-security systems. Today's advocates of space jets, lunar vacations, and the manned exploration of the solar system appear to hail from another planet. A faded 1964 Popular Science cover story — “Who'll Fly You at 2,000 m.p.h.?” — barely recalls the dreams of a bygone age.

— Peter Thiel, *The End of the Future*

Advances since 1970 have tended to be channeled into a narrow sphere of human activity having to do with entertainment, communications, and the collection and processing of information. For the rest of what humans care about—food, clothing, shelter, transportation, health, and working conditions both inside and

outside the home—progress slowed down after 1970.

— Robert Gordon, *The Rise and Fall of American Growth*

If one sentence were to sum up the mechanism driving the Great Stagnation, it is this: Recent and current innovation is more geared to private goods than to public goods. That simple observation ties together the three major macroeconomic events of our time: growing income inequality, stagnant median income, and the financial crisis.

— Tyler Cowen, *The Great Stagnation*

Regardless of the mechanism, it's clear that *something* has changed about how we as a society create new knowledge and technology. I don't think we can ever know with certainty why the Great Stagnation exists, what will end it, or whether it even existed in the first place. However, I do know that *we can do better*.

These kinds of examples show that there *can* be ecosystems that are better at generating progress than others, perhaps by orders of magnitude.

— Tyler Cowen and Patrick Collison, *We Need a New Science of Progress*

It's time for full-throated, unapologetic, uncompromised political support from the right for aggressive investment in new products, in new industries, in new factories, in new science, in big leaps forward.

— Mark Andreessen, *It's Time to Build*

One way to unleash our potential is to build new institutions. New institutions can enable activities that current institutions do not. Weirder experiments to unlock new forms of matter. More ambitious projects to push nature to its limits. Sitting with the unresolvable uncertainty that stands in the way of greatness. Curiosity coupled to unresolvable uncertainty. Unless you believe everything is fine, enabling currently unsupported activities is essential to build a future that

maximizes human capability and wonder.

It's easy to point out flaws in existing institutions, but hard to craft new ones. One way to cheat (the deck is stacked against us so we need every advantage we can get!) is to study the paragons of the past. Some names leap to mind: Bell Labs, (D)ARPA, Xerox PARC. It's tempting to either copy them directly or gesture vaguely towards them as 'inspiration.' Instead, our goal should be to deeply understand why they worked and then riff on that understanding for today's world.

DARPA is the most promising paragon to riff on. This assertion is counterintuitive — DARPA is a government-run military research organization; when you think of institutions that built the future, Bell Labs and Xerox PARC dominate. However, those institution's success depended on a set of conditions that do not exist today. On the other hand, private organizations riffing on DARPA's model are not impossible. It may even be possible for them to create the outlier results that shift an incredible future from a possibility to an inevitability.

[An industrial lab] is merely an organization of intelligent men [sic], presumably of creative capacity, specially trained in a knowledge of the things and methods of science, and provided with the facilities and wherewithal to study and develop the particular industry with which they are associated.... It is likewise an instrument which can bring to bear an aggregate of creative force on any particular problem which is infinitely greater than any force which can be conceived of as residing in the intellectual capacity of an individual.

— Frank Jewitt, First President of Bell Labs, *The Idea Factory*

While science generates much of our prosperity, scientists and researchers themselves do not sufficiently obsess over how it should be organized.

— Tyler Cowen and Patrick Collison, *We Need a New Science of Progress*

Solutions R&D weaves the threads of research from multiple domains together

with lessons from the reality of use and practice, to demonstrate prototypes, develop tools, and build convincing evidence. Because it reaches into and connects all the parts of the innovation system, solutions R&D is a powerful way to ratchet the whole system up faster, once some initial elements of research and implementation are in place. Doing it well takes a management approach that combines a relentless focus on a bold goal with the ability to manage the high risk involved in creative experimentation.

— Former DARPA Director Arati Prabhakar, *In the Realm of the Barely Feasible*

There is an entire maze of decisions that an organization riffing on the DARPA model needs to navigate. Legal structures; how money works; incentives; how you find and work with world-class researchers in the first place. On top of that, there's not just one path through the maze! In this piece I'll sketch the edges of the idea maze and point out the deadly traps and Minotaurs that lie within; I'll describe the contours and forks of the maze and lay out my plan to navigate it. By the end, I plan to convince you that the plan is not entirely half-baked insanity and hopefully equip any other adventurers who also want to riff on the DARPA model or build new research institutions.

I cannot help fearing that men may reach a point where they look on every new theory as a danger, every innovation as a toilsome trouble, every social advance as a first step toward revolution, and that they may absolutely refuse to move at all for fear of being carried off their feet. The prospect really does frighten me that they may finally become so engrossed in a cowardly love of immediate pleasures that their interest in their own future and in that of their descendants may vanish, and that they will prefer tamely to follow the course of their destiny rather than make a sudden energetic effort necessary to set things right.

— Alexis de Tocqueville

Purpose

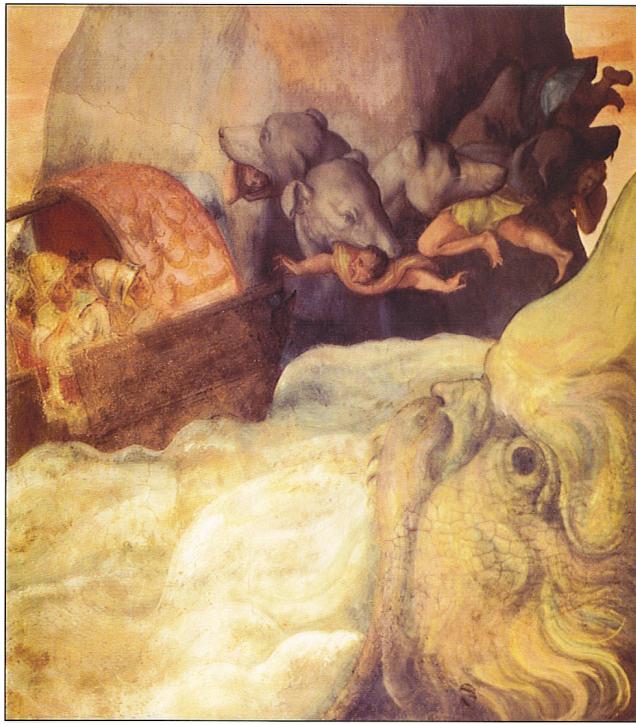
This piece is a combination of three components: a specific organization's design; a broad proposal for a new organizational structure and an associated 'research' agenda; and a synthesis of a gestalt of institutions, incentives, history, and theory that stabs at the questions "how does the process of creating new impactful knowledge/technology work?" and "how can we do more of that better?"

It's traditional to separate the three roles this document is trying to play — analytical synthesis, broad proposal, and specific organizational design. However, in this case they are too intimately entangled. Separating any one of them would force me to either leave glaring gaps or project an unwarranted level of certainty. The synthesis clearly drives the broad proposal as well as specific organizational hypotheses. A large part of the research agenda is inseparable from the specific organizational design itself — from "how does money work in a DARPA-riff?" to the hypothesis that good simulations could be the 'why now' of a program design discipline. At the same time, one of the (admittedly grand) goals of the specific organization is to shift some of the conclusions in the synthesis; coupling the synthesis to the other parts allows us to say not just "here is how the world is" but also "here is how it could be."

An important goal for this piece is to argue for the existence of a new organization in the first place. Creating an organizational structure around a project prematurely can have many downsides, so the burden of proof rests heavily on the argument that you can't accomplish what you need to do without an 'official' organizational structure. I will argue that there are a number of experiments that seem like they can only be done in the context of a new organization and that they comprise some of the most interesting questions that the broad proposal wants to explore.

Calling out and exploring unavoidable tensions is one of the piece's core themes and an important role. Innovation organizations are shot through with Straights of Messinas, with Scylla waiting to snatch you if you focus too far on side of a tradeoff and Charybdis ready to suck you to your doom if tack too far the other

way. As we will see, these tensions often have to do with incentives. With these tensions, the only way to avoid a messy fate is to first know the upcoming danger and then to sail the narrow path between them, which will always be uncomfortable and require constant course correction. These tensions manifest even in the piece's very existence: if you think you have good ideas, you should show that by acting on them; at the same time, propagating the ideas can also be valuable if it enables other people to act on them. But acting on ideas and explaining them are often at odds! These conflicting prerogatives are especially true in this context where we need an explosion of new models, not a single organization. The piece's ultimate goal is to affect change, but the nature of the beast is such that I suspect neither generalizable knowledge nor a specific organizational design will be sufficient on their own. Normally people end up on one end of the spectrum or another — either publishing a policy proposal or trying to lead by example (which doesn't even produce an artifact).



"Om nom nom" — Scylla and Charybdis

Finally, this piece is meant to build trust. Doing anything new requires trust and research requires more trust than other disciplines. Many of these ideas don't

have a ‘closed form’ solution or a right answer, so the only way that I can convince you that I’ve come to a reasonable conclusion is by walking you through the process of getting there. Hopefully I can build that trust by showing you that I’ve really ‘done my homework’ and walking you through the thoughts behind the actions. There are many uncomfortable truths and icky tradeoffs embedded in good solutions R&D. Staring them in the eye requires trust in addition to raw logic.

Who is this piece for?

This piece is intended for five types of readers. I’m explicitly calling them out because the whole thing is long and everybody will get more or less value from different sections. You don’t need to be a completionist!

1. **“The builder”**: You’re this type of reader if you want to put in the legwork to build an innovation organization, especially one that riffs on the DARPA model. In other words, me. In some ways this piece is a very long memo to myself laying out how to spend the next N years of my life and why I’m doing it. If you’re in this category, you already (at least slightly) believe strongly enough in the importance of creating new innovation organizations that you want to take action on that belief. Builders are probably most interested in the general principles for building DARPA-riffs in Part II and less interested in the more abstract arguments in Part I and the specific plans in Part III.

2. **“The supporter”**: You’re this type of reader if you’re in the position to support new innovation organizations — monetarily or otherwise. Creating new innovation organizations isn’t going to be your main thing, but you still would like to put your social or financial capital towards the effort. You could be coming from one of many institutional contexts: as a philanthropist, a member of a government, or a capitalist — either an investor seeking monetary returns or the representative of a corporation looking for technological returns. Supporters are probably most interested in the justifications for building DARPA-riffs, the

discussions around money and profit, and specific plans.

3. **“The enthusiast”**: You’re this type of reader if you are generally interested in the question “how do we enable more creative science and technology in the world?” but perhaps you aren’t in a position to be a builder or supporter … yet. Enthusiasts are probably most interested in the history, big conclusions, and possibilities for further exploration in Part I and II along with program ideas in Part III.

4. **“The scholar”**: You’re this type of reader if you have an academic interest in innovation and how it works. Maybe you’re a historian, economist, or policy person. Scholars are probably most interested in the history and theoretical discussions in Part I along with bigger conclusions in Part II.

5. **“The observer”**: You’re this type of reader if you’re a curious person who doesn’t fall into any of the previous three categories but enjoy learning about the world and coming across new ways of thinking. Perhaps you have a professional role where you benefit from understanding what’s going in the world of research and emerging technology. Perhaps not!

Of course, you can always **chart your own adventure**.

Housekeeping

Links. Whenever I lean on an argument I make elsewhere in the piece, I try to use internal links that look like this (which will take you right back to this section). You shouldn’t worry about losing your place because clicking an internal link will make a link back to the section you jumped from appear in the upper right. External links [look like this one that actually links back to this piece](#). I try to only use them when it’s painfully obvious what they link to, usually the name of a piece I’m referencing directly in the text.

Guided tours. This piece is long and everybody will get more or less value from

different sections. Selecting a persona in the section above will highlight the headings of some sections in the table of contents and grey-out the headings of others in a ‘recommended path’ for that persona. It won’t make anything disappear. You can get rid of the highlighting by choosing “chart your own adventure.”

Money Notations. When I refer to historical monetary values, I’ve converted them into 2021 dollars as indicated by eg. \$(2021)1B for \$1B in 2021 dollars.

Foot and sidenotes. This piece has both sidenotes and pop-up footnotes. Sidenotes are for external references to expand on a topic while footnotes are for extraneous asides — you will lose nothing by not reading them.

Playfulness. This piece deals with a serious subject (to me it is one of *the* most serious subjects). It’s easy and often expected to take a grim authoritative tone to convey seriousness. Yet, examples from Einstein to Feynman plates to Shannon suggest that playfulness not only does not get in the way of good work, it may be an ingredient. With that in mind, I’ve allowed myself to be a bit playful and irreverent at times and hope that doesn’t diminish the piece’s seriousness in your eyes!

Institutional Design is Navigating an Idea Maze

Navigating an “Idea Maze”⁵ is a powerful analogy for a process that involves a series of hard-to-reverse decisions under uncertainty. Idea Mazes usually show up in the context of startups and the decisions an entrepreneur navigates to create a company — “should we sell to businesses or consumers? Build a web application or iPhone App first? etc.” Despite being associated with startups, there is nothing about the analogy of an idea maze that restricts it to a particular domain; it can also be used in engineering, research, or in this case, building new

⁵ As far as I can tell, Balaji Srivasanan introduced the Idea Maze in [his Stanford Startup Course](#) and Chris Dixon made it well known [on his blog](#).

institutional structures.

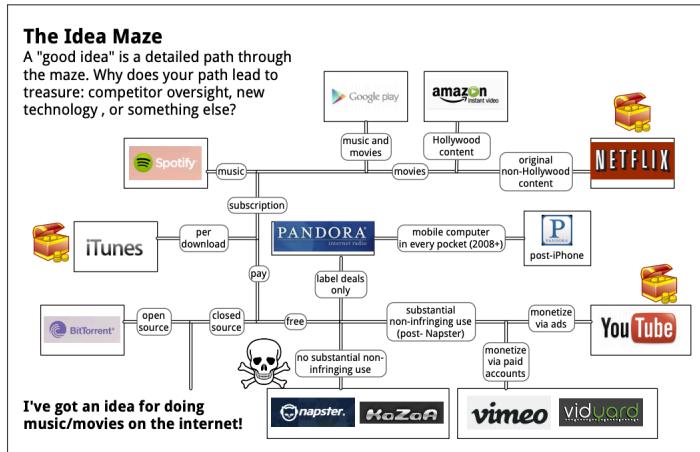


Image Credit: [Balaji S. Srinivasan](#)

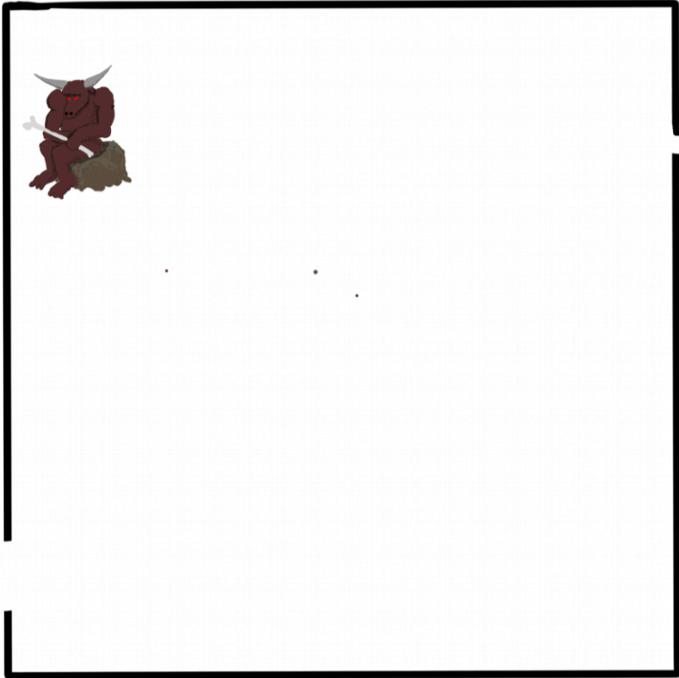
Both abstract and concrete ideas are essential for talking about institutional structures; extending the analogy of the idea maze to the full myth of the Cretan Labyrinth can act as intellectual glue between the two. On the one hand, it's important to talk about *why* a new structure is even necessary and sweeping considerations that apply regardless of specifics. This is like the looking at the maze from the outside — where are its boundaries? Why does it exist? What dangers lurk within? At the same time, it's equally important to talk about gritty implementation details. These are the junctures in the maze and how you know that you've reached them. The abstract considerations are important for enabling people to come to their own conclusions but without the details, it's easy to leave the difficult tradeoffs as frustrating 'exercises for the reader.'



No not that labyrinth.

This piece has a three-part structure that addresses increasingly concrete questions. The first question it strives to answer is “Why does the maze exist and where are its boundaries?” The second question is “What is the layout of the maze?” And finally, “What is the specific path we will try to chart through the maze?”

Why does the maze exist and where are its boundaries?

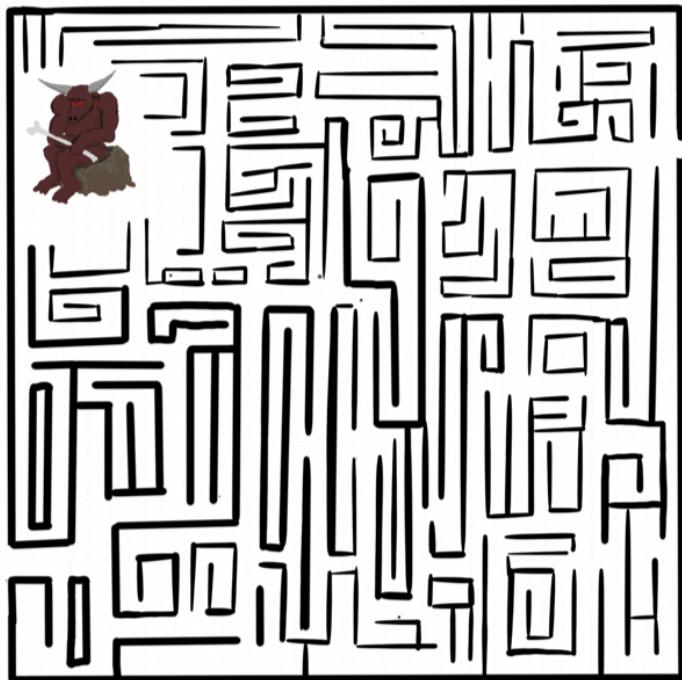


When your goal is to navigate a maze, it's easy to treat it as a fixture in the world and ignore what it looks like from the outside and its history. But mazes exist for reasons. Don't forget that King Minos had the Labyrinth built in order to contain the minotaur! Describing the boundaries of the maze in the context of institution building is about understanding the scope of what we're setting out to do and the constraints in the world that shape that scope. The history lets us know what's lurking inside to kill us and eat us.

The exterior of the maze is fractal — every time you zoom in on it, there are more interesting details to explore. As such, it's easy to fall into the trap of focusing *only* on the contours of the questions “why does this problem exist?” or just “is there is a problem?” Answering these is a worthy goal but it is not our goal. I spend a good chunk of the piece describing the exterior of the maze and why it exists, but only insofar as I feel it's useful for hypothesizing about the maze's interior and the path through it. This approach will frustrate some of you for its lack of rigor and others of you for extraneous information. Not all sections

are for everybody!

What is the layout of the maze?



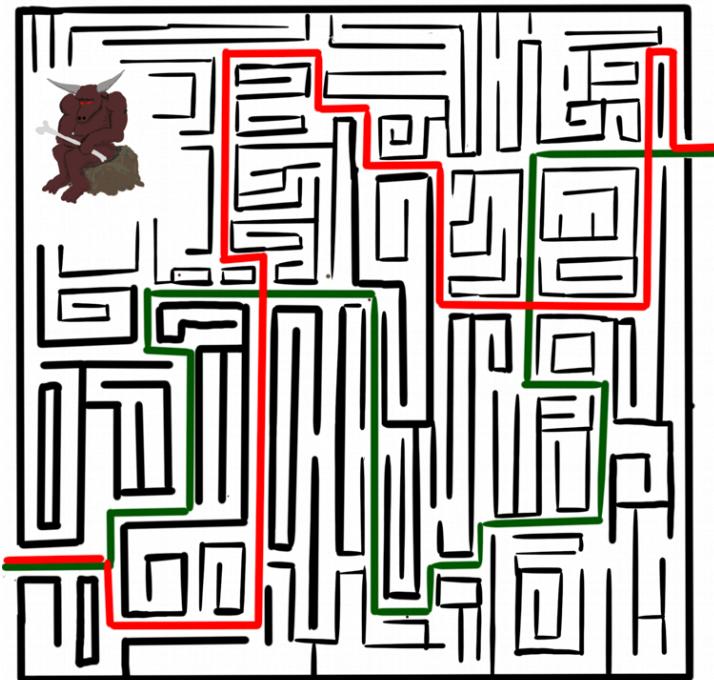
The interior layout of the maze corresponds to the myriad decisions facing anyone who is trying to build an organization to tackle a specific niche. The layout of the maze — with its junctures, turns, and dead-ends — is impossible to know for sure before you enter it. Even then you'll only be sure of the junctures and distances you've actually traversed and only if you keep careful track of them. However, it is possible to sketch possible layouts. This sort of hypothesizing important and underdone, especially in public, in large part because there is not just one path through the maze! In addition to explaining which decisions I think are correct, I want to both encourage other adventurers to enter the maze and equip them as well as I can.

It's also just good practice to lay out your hypotheses about not just what experimental results will be but *why* they will happen. Abstractly describing decision points is admittedly a bit of a hedge! Even if our specific implementation fails, I hope that it doesn't stand as a condemnation of the entire

model.

I'm going to spend a good chunk of the piece describing the potential junctures that someone building a DARPA-riff will probably face and how to think about them *before* explaining which fork I think is correct.

What is the path that we hope to chart through the maze?



Finally, there is the actual path through the maze. Any organization can only follow a single golden thread, making one specific design decision at each juncture. I'll lay out which choices I think are correct given the picture of the maze I've constructed. This is the most concrete part and also the most likely to be wrong or need to change. It has been nerve-wracking to write for two reasons. First, concretely stating "I think doing this will work" opens you to the terrifying possibility of being provably wrong. When you make general statements about the world, you can always point to qualifiers about why an example does not count as a disproof. Not so when you say "I am going to try to do X and expect Y to happen." The second reason that laying out a specific path is nerve-wracking is because people tend to evaluate ideas on their most concrete manifestations. So

in the same way people accept entire ideologies on a few compelling anecdotes, they are likely to reject the entire path or possibly the entire argument I've laid out based on one dumb choice.

For some readers this will be the most interesting part — concrete, actionable plans. For others this will be the most boring.

Different disciplines usually focus on one part of the maze at a time. Historical and economic work usually focuses only on why the maze exists and its boundaries. Policy proposals usually focus on the layout of the maze.

Entrepreneurs usually focus on specific paths through the maze (but don't write about it except in retrospect). Chesterton's fence⁶ suggests that focusing on one part at a time is probably good practice but in this specific situation it feels like each piece would be weak without the others. Describing the state of the world would just be adding to a growing pile of stagnation literature if it wasn't supporting action. Without the context provided by the boundaries, the maze's layout would be hard to evaluate, and without the specific plan it would feel like yet another call for someone to do something! And without the boundaries and layout, it would be impossible to evaluate a proposed path beyond "well that seems smart/stupid" or have a discussion about how it could be improved from a common framework.

⁶ Generally, things are done the way they are for a reason.

Problems and Possibilities



In this section I want to convince you that

1. In the past, industrial Labs filled a particular niche in the innovation ecosystem.
2. Especially in the world of atoms, industrial Labs no longer fill the niche they did in the early-mid twentieth century.
3. The niche industrial Labs once occupied still needs to be filled and we should not expect organizations that look like industrial labs to fill it.

Overlapping institutional constraints rule out several classes of creative work

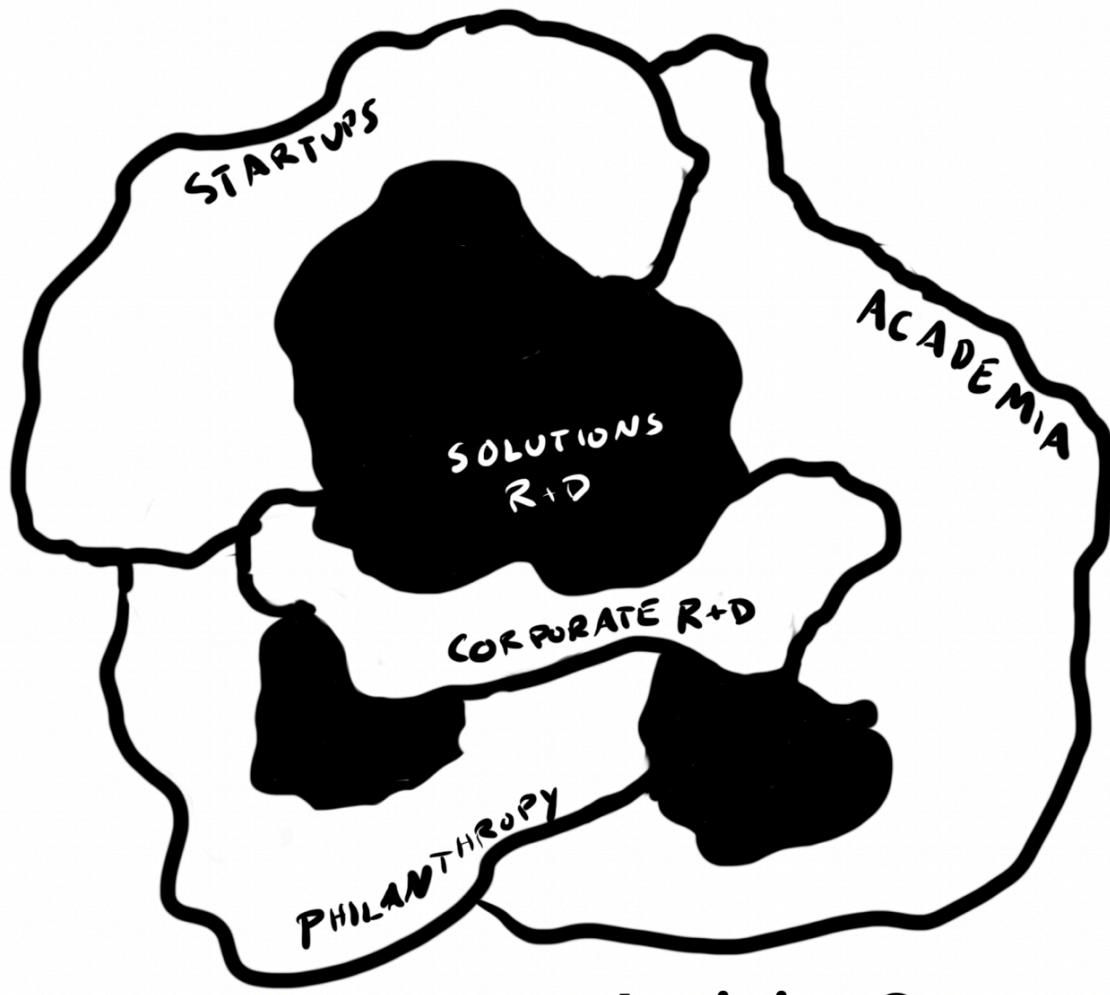
Different institutions are each good at enabling a certain set of activities that we associate with ‘innovations’: academia is good at generating novel ideas; startups are great at pushing high-potential products into new markets; corporate R&D is unparalleled for improving existing product lines. Together, the institutions that play a part in creating new knowledge and technology comprise an ‘innovation

ecosystem.⁷ The organizations that make up these institutions vary wildly, but they share enough commonalities that it's worth collectively referring to them as 'innovation organizations.'

Obviously, each institution can't excel at every type of activity. Each institutional structure has some set of constraints that prevent it from engaging in certain activities well or at all. Many factors shape these constraints and they are effectively synonymous with 'incentives.' You could think of each institution occupying some area on a map of all possible activities — the institution is well suited to tackle the activities it covers and poorly suited to tackle activities at the edge or outside of its area. Some activities are covered by multiple institutions, but some aren't covered by *any* institutions. These activities are 'constrained out' of happening by existing institutional structures. For example, unsexy long-term⁸ projects that won't necessarily produce a product or novel papers are simultaneously beyond the scope of venture-funded startups, corporate R&D, philanthropy, and academia. Venture funded startups could support either a shorter timeline *or* more product focus; corporate R&D would want the activities to be sexy or product-focused; philanthropy would want them to be sexy; and academia would push for novelty and papers.

⁷ Both the terms 'innovation' and 'innovation ecosystem' are often frustrating suitcase words. But it is useful to have a shorthand for "all the institutions that enable activities that collectively cover the process of an idea becoming a new impactful idea or technology."

⁸ Long-term roughly translates to more than ~five years.



Of course there are counterexamples of projects with those characteristics that have been supported by existing institutional structures! It will always be possible to make arguments like “SpaceX exists, so the innovation ecosystem is fine.” Antibiotics are pretty great, even though people can sometimes survive infections without them. Similarly, it’s worthwhile to try to enable more constrained activities even if a few projects make it through.

Institutionally constrained activity is a useful and precise way to think about the vibe that the world could be on a more wonderful trajectory than it’s on right

now.⁹

Instead of arguing over The Great Stagnation¹⁰ or lamenting the fact that some combination of research/academia/science/physical innovation is broken, looking through the lens of institutional constraints enables us talk specifically about which classes of activities we expect to be happening that are missing or anemic. We can then ask what institutional constraints are creating those gaps. Instead of asking “where’s my flying car?”¹¹ we should be asking “what activities would enable flying cars and what institutional constraints are stopping them?”¹² This analysis can in turn help us address those gaps by pointing to concrete ways to adjust incentives in existing institutions or which new institutions with different sets of incentives one could create. Some important questions to ask are “what incentives are preventing institutions from enabling these activities? Should I be shifting incentives within an existing institution or building a new one? How do I keep a new institution from falling into the same incentive traps?”

Looking at the innovation ecosystem through this lens suggests that there are many different activities that we lump together under the umbrella of ‘research’ or ‘innovation’ and many ways they could be improved. There are infinitely divisible multitudes, but I will briefly note three clusters of them primarily because they are frequently conflated and I explicitly want to say “this piece is about one of these but not the other.”

One cluster where things could be better is what I might call “breakdowns in the scientific process.” Here you see issues like the replication crisis and science being judged not on one of several scientific epistemologies, but on politics and

⁹ My unsubstantiated personal opinion is that analyses on high-level metrics like total factor productivity contribute to this unhelpful lumping.

¹⁰ See *The Great Stagnation: How America Ate All The Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better* and a lot of related literature.

¹¹ See *Where is My Flying Car?: A Memoir of Future Past*

¹² Yes, it’s a much less catchy question. And in this particular case the book does delve into the institutional constraints; many people do not.

its ilk.¹³

Another area for improvement is in enabling paradigm-shifting work. It's a bit trope-y but there is something in the fact that someone from 1920 would barely recognize the built world in 1970, but someone from 1970 wouldn't be too surprised by our cars, planes, and buildings or their capabilities. The built world hasn't experienced many paradigm shifts. Similarly, physicists are still working on string theory fifty years later and despite incredible advances in the science biology, nothing has displaced the structure and centrality of DNA as the dominant biological paradigm. There are legitimate arguments that this narrative ignores less-visible paradigm shifts like those in computing or the fact that the late 19th -> mid 20th century may have been a massive outlier in the realm of physical innovation. I don't think these arguments are incompatible with the assertion that we could do better.

You can roughly divide paradigm shift-enabling activities into two distinct modes. *The first mode is paradigm-shifting science.* Here the sentiment is that "Einstein would be stuck in the patent office." We don't seem to have a healthy system of unfettered research that is producing paradigm shifts in how we understand the world like those seen up through the 50's or 60's in everything from quantum chromodynamics to general relativity to the double-helix structure of DNA. *The second mode is paradigm-shifting engineering.* "Where's my flying car?" We don't seem to be able to build new systems that radically change our physical capabilities. The line between science and engineering is of course nebulous, porous, and full of feedback loops.

Distinguishing between these areas isn't just a semantic exercise — it's important for enabling action. While they share many similarities and causal links, each area prioritizes different activities and mindsets. As a naïve example, paradigm-shifting science probably depends on monomaniacal individuals running an idea to ground over the course of sometimes decades while that same mode would be

¹³ For more on breakdowns on the Scientific Process, I recommend Brian Nosek's work on the replication crisis or Stuart Ritchie's book *Science Fictions*

detrimental to paradigm-shifting engineering because it requires more pragmatism and coordination. Think Einstein and general relativity vs. Polavision — Polaroid's failed home movie system.¹⁴ It's important to address each of these areas but there needs to be a division of labor. Despite the temptation to do all the things, 'fixing research' is going to require *many* new institutions (that I suspect will not scale) to address different niches in constraint space. Additionally, legible niches can help prevent mimetic infighting over who is going to 'save science.' Given all of this, enabling more paradigm-shifting science is *incredibly important* but it won't show up in the rest of this piece as we turn our focus to the niche in the innovation ecosystem that enables paradigm-shifting engineering.

Feeling Around the Edges of Solutions R&D

How should we think about the bundle of activities that enable paradigm-shifting engineering? The bundle is incredibly nebulous — absolutely real, but hard to capture, fuzzily edged, and context-dependent. There is clearly a *something* that is more focused on products and outcomes than the exploration of nature we associate with Newton or Einstein yet at the same time doesn't have the steely-eyed focus on commercial products we associate with Edison or Jobs. Invoking these contrasting historical figures brings to mind Donald Stike's concept of Pasteur's Quadrant¹⁵; Stike captures the mindset excellently but doesn't focus on the actual institutional activities or how to enable them. Closer to the mark, DARPA Director Arathi Prabhakar captures the spirit of what we're striving for in her description of "Solutions R&D":

Solutions R&D weaves the threads of research from multiple domains together with lessons from the reality of use and practice, to demonstrate prototypes,

¹⁴ Polavision took ten years and half a billion dollars. In *Loonshots* Safi Bahcall attributes its failure to expensive but amazing features driven by Polaroid's founder.

¹⁵ See Pasteur's Quadrant: Basic Science and Technological Innovation

develop tools, and build convincing evidence. Because it reaches into and connects all the parts of the innovation system, solutions R&D is a powerful way to ratchet the whole system up faster, once some initial elements of research and implementation are in place. Doing it well takes a management approach that combines a relentless focus on a bold goal with the ability to manage the high risk involved in creative experimentation.¹⁶

“Solutions R&D” is certainly a great name for the niche we’re interested in and important — we don’t have to keep referring it to as “that bundle of activities that enables paradigm-shifting engineering” (though it lead to an impressive word count.) But like other nebulous things such as clouds, Science, and porn a name doesn’t do much to further understanding the important properties of the activities we’re talking about, why they’re not as common as we’d like, and how to change that.

Perhaps frustratingly, the best strategy for getting a handle on nebulous but very real things is not to rigidly define them but to “feel around the edges.” People are terrible at describing what they want given a blank slate, but we’re pretty good at knowing what we want when we see it. Considering what triggers the “yes, that!” for systems R&D, we consistently arrive back at the industrial labs of the early-mid 20th century and Bell Labs in particular.

In the past, industrial Labs filled a particular niche in the innovation ecosystem

Solutions R&D is nebulous, but we can get a sense of it by feeling around the edges and looking at what sorts of things happened at great labs hat don’t seem to be happening now. While industrial Labs (Including Bell Labs and PARC themselves) still exist, industrial Labs no longer fill the niche they did in the early-mid twentieth century. Understanding what changed and why is part of understanding the niche more broadly. While the early-mid 20th century was

¹⁶ From In the Realm of the Barely Feasible

home to many well-regarded industrial labs¹⁷ like GE Laboratories (f. 1900), DuPont Experimental Station (f. 1903), Kodak Research Laboratories (f. 1912), Bell Labs (f. 1925) stands head and shoulders above them in the pantheon. Most of the properties and important activities in this section are drawn primarily from accounts of Bell Labs, supplemented by accounts from other labs. This Bell Labs focus is admittedly in part because of the streetlight effect — people have chronicled Bell Labs much more extensively — but also because Bell Labs' output was such an outlier. I feel comfortable doing this because unlike the idiosyncratic success of individuals or startups, there appears to be a consistent set of activities shared across industrial labs that are missing today.

The trick here is to distinguish between the ‘universal’ characteristics of the activities that enables good solutions R&D and the particular set of tactics and strategies that Bell Labs used to implement those characteristics. It’s important to distinguish between the two because we might need to use different tactics to achieve the same characteristics in a new 21st century context. Blindly following tactics regardless of context is ineffective and cargo-culty.

I’ve identified **nine characteristics that enabled industrial labs to fill the solutions R&D niche.**

1. Industrial labs enabled work on a general purpose technology before it’s specialized.
2. Industrial labs enabled targeted piddling around, especially with equipment and collaborators who would not otherwise be accessible.
3. Industrial Labs enabled high-collaboration research work among larger and more diverse groups of people than in academia or startups.
4. Industrial Labs enabled smooth transitions of technologies between different readiness levels — they cared about both about novelty *and* scale.

¹⁷ For more about why industrial labs started see The Changing Structure of American Innovation - Some Cautionary Remarks for Economic Growth

5. Industrial Labs provided a default customer for process improvements and default scale for products.
6. Industrial Labs often provided a precise set of problems and feedback loops about whether solutions actually solved those problems.
7. Industrial Labs provided a first-class alternative to academia where people could still participate in the scientific enterprise.
8. Industrial Labs enabled continuous work on projects over a 6+ year timescales. The work to create the transistor took eight years (because of a gap created by WWII) and making it mass manufacturable took another eight.
9. Industrial Labs enabled work in Pasteur's Quadrant¹⁸ — explorations into natural phenomena with eye towards exploiting those phenomena for human flourishing (and profit).

Exceptions abound of course. Non-industrial lab organizations exist today that embody many of these characteristics and not all golden-age industrial labs did all these things all the time. I'll expand on some (but not all) of these points below.

Industrial labs enabled work on a general purpose technology before it's specialized

Specialization has a complicated relationship with new technologies. New technology will never be as good as old technology along every dimension.¹⁹ This failure of pareto-superiority is why frontier technologies need to start in niche markets where they are especially valuable. It's rare that a technology can do well in a niche market without *some* specialization. If you

¹⁸ Pasteur's Quadrant is an excellent exploration into the nuances of different types of research.

¹⁹ Invoking Cunningham's Law, I will say this very aggressively in the hopes that I am wrong.

imagine the straightest path that a technology can take to its most powerful or general use,²⁰ the specialization work to fit into niches can require larger or smaller diversions from this path. These diversions take the form of everything from technology development to marketing to building sales channels for a company. Going from niche to niche is essential, but you can also get stuck in a niche or jump onto a different development path altogether. Say what you will about him, but Elon Musk is the master of charting paths between niches: roadster -> high-end sedan -> mass market car; NASA-subsidized LEO rocket -> GEO rocket -> reusable GEO rocket -> Starship. More time and resources can enable you to pick an “optimal” sequence of niches or search around for yet-unknown niches instead of being forced to hop into the nearest or most obvious ones.²¹ In other words, if you imagine decoupling from market discipline as cave diving, industrial labs acted like extra oxygen tanks.

Industrial labs’ work on unabashedly general purpose technologies stands in contrast to modern startups. I don’t know how often I’ve seen a presentation from a grad student or professor that goes something like “we’ve invented this incredible technology that could potentially do <amazing thing> however it will take a lot of work to get there, so we’re starting a company to build a product to do <kind-of-pathetic thing> that will still require a lot of specialization and company-building> in <domain totally unrelated to <amazing thing>>.” While I’m usually sure it’s doomed to failure, it’s unfortunately the best move given history and the constraints on startups. First, many successful startups *do* look like they started in a tiny scoffable niche — Tesla (weird rich people), NVIDIA (gamers), PayPal (Beanie Babies on eBay?), Apple (hobbyists), Twitch (life live-streaming), Amazon (books) … the list goes on. However, each of these weird niches didn’t require huge diversions from the critical path.²² Second, ‘start in a

²⁰ This idea might be a fallacy. Arguably there is no platonic “ultimate stage” of a technology and whatever a technology evolves into is indelibly touched by the different specialized forms it needed to pass through.

²¹ More abstractly, general purpose technology needs developmental *slack*. See Studies on Slack.

²² There are a number of just-so stories I could tell about why these university spin-offs are different than the list of successes - most of those on the list were not deeply new

'niche' is folk startup wisdom *for a reason*. History is littered with the corpses of startups that were going to be the next platform or general purpose technology as soon as they launched. General Magic, NeXT, Quibi, Atrium, Rethink Robotics, and Magic Leap. Many others that started with grand ambitions were forced to crawl into a niche to survive - I won't name names here. At the same time, startup investors want to see *massive* potential returns. 'Find a compelling story about how a niche becomes a billion dollar market' puts a technology between a rock and a hard place.²³

In most cases, it wouldn't have been feasible to do the necessary work to find or develop for a better niche in an academic setting either. While that work still requires "research" it also requires focus and systems engineering, both things that academia does not support for a slew of reasons.

Of course, it's possible that I and others just overestimated the potential of many of these technologies. However, the examples of technologies that were massively impactful only after 'hanging out' in an industrial lab for many years would seem to argue otherwise. The transistor, public key cryptography, pyrex, solar panels, and the Graphical User Interface (and personal computing in general) come to mind. Industrial labs provided the environment to do the work to get a technology to a point where it's viable for a 'useful' niche and enable it to find that niche in the first place.

In addition to giving projects longer timescales, less existential risk, and larger budgets than most startups, there are several specific ways that industrial labs helped technologies find good niches that startups and academia don't provide.

When you're still trying to figure a technology out, it's not clear which skillsets

technology, their founders were brilliant and plotted the exact path to success, it's easier to go into new niches with software and chips, they got lucky, I'm a moron etc. However, the empirical evidence is that so many niche-seeking startups built around new (genuinely cool) technologies fail to help that technology realize its potential.

²³ I'm not trying to say VC and startups suck and are broken! There are many amazing things that have come out of the system. What I am asserting is the incorrectness of the idea that "if it wasn't a viable startup it would never have been able to be amazing for the world in the long run"

you want in the room. Industrial labs facilitated people floating between different projects loosely creating and breaking collaborations. Bell Labs was particularly good at enabling these free radicals:

The Solar cell just sort of happened,” he [Cal Fuller] said. It was not “team research” in the traditional sense, but it was made possible “because the Labs policy did not require us to get the permission of our bosses to cooperate—at the Laboratories one could go directly to the person who could help.²⁴

To get the same effect at a startup, you need to either find a magical individual who has gone deep on multiple areas (an M- instead of T-shaped individual?) or hire people who might ultimately turn out to be useless. Startups don’t have the slack for this.

Another dividend from industrial labs’ slack is the room to notice something unexpected, say “huh, that’s funny,” and run the anomaly to ground. Perhaps the most famous slack-dividend is the discovery of the cosmic microwave background when Arno Penzias and Robert Woodrow Wilson were using the Holmdel Horn Antenna to try to do satellite communication experiments. They noticed persistent noise in their measurements that didn’t seem to respond to recalibration or cleaning. Eventually they got in touch with theorists who connected the noise they saw with predicted echoes of the (still very hypothetical) big bang. In a lower-slack (more ‘efficient’) organization, Penzias and Wilson wouldn’t have had the bandwidth to dig into the mysterious noise beyond determining that it didn’t seem to be fixable.

Modern corporate R&D has become worse at enabling work on general purpose technology. Bell Labs had a culture that was almost the polar opposite of the ‘demo or die’ pressures in many modern organizations. The need to produce demos for corporate higher-ups may create the same pressure to specialize into a niche that makes startups a bad place to develop general purpose technology. One way to look at this would be to ask ‘how has the *management* of corporate

²⁴ From The Idea Factory

R&D changed over time?

Industrial labs enabled targeted piddling around

So many stories of golden-age industrial labs revolve around researchers literally just trying out a lot of stuff. “Hey, we came up with this new chemical, what’s it useful for?” In the startup world, that’s often derided as a “solution looking for a problem.” Arguably though, many important technologies started off this way — they weren’t a hole-in-one solution to a problem. We don’t seem to have a place for this targeted piddling around²⁵ to happen anymore for situations that require specialized knowledge and equipment.

Disciplines where targeted piddling around can happen seem to be healthier. This contrast is stark looking at software vs. say, space technology. Modern software (and slowly, biology) suggests that in many cases, piddling around requires cheap, democratized technology. Industrial labs may have been able to loosen that requirement. Stories about Dupont and 3M’s industrial labs always involve a lot of just ‘trying stuff out.’

It’s important to call attention to the ‘targeted’ part of ‘targeted piddling around.’ Contrary to common perception, Bell Labs and PARC *didn’t* give researchers free rein to work on whatever they wanted. There were relatively few milestones and researchers had enough slack from management to explore adjacencies.

However, people were explicitly asked to work on high level goals that would benefit AT&T’s continent-spanning communication system.

This targeting stands in contrast to accounts extolling how much freedom researchers had. It’s almost impossible to confirm, but I suspect that researchers who portray themselves as having completely free rein in industrial labs are

²⁵ This is a made-up term inspired by Surely You’re Joking Mr. Feynman: “It was effortless. It was easy to play with these things. It was like uncorking a bottle: Everything flowed out effortlessly. I almost tried to resist it! There was no importance to what I was doing, but ultimately there was. The diagrams and the whole business that I got the Nobel Prize for came from that piddling around with the wobbling plate”

unreliable narrators. My hunch is that they felt like they had entirely free rein because the lab managers were good at hiring people whose interests were sufficiently aligned with the lab's goals that anything they chose to do was within some window. Another alternative is that they gave a few people free rein and the personal accounts are from the wunderkinds. Claude Shannon illustrates both of these possibilities — he was absolutely a wunderkind who was given free rein. That is, up until he started working on projects like a 'mechanical mouse' that Bell Labs management couldn't possibly justify to regulators as being related to communication.

Today, many people implicitly assume that targeted piddling around happens in universities. Academics certainly do piddle around with *ideas* but they're less incentivized to do it with technology applications. You can't really write a paper about how you spent a year modifying your novel technology to 100 specific applications. Even when academic labs do piddle around with technology applications, the targeting systems are often suboptimal. Often academics are working on a specific application for a specific industrial partner or they've effectively made up a use-case out of whole cloth. This isn't to cast shade on academics but to point out that academia is not set up to enable the sorts of feedback loops to enable just the right amount of targeting.

Getting a general purpose technology work is actually an iterated cycle of generality and specificity. You can think of two coupled modes: testing a general tool on many specific applications and testing many specific components to get a general tool to work. Xerox PARC's GUI work is a good example of the former mode — they stress tested the general system by using it for day-to-day work! Edison's Workshop (arguably the OG industrial lab) did a lot of 'trying stuff out' but it was more the opposite way, where they had a general application in mind (lighting, recording, picking up voices on the phone) and they just went through a thousand materials to see if they would fit the bill.

The two modes are strongly coupled. While you're trying out a thousand things to do a specific thing, you may realize that the five hundred and twenty first thing is no good for the intended purpose, but might be amazing at another

thing. However, that realization can easily be tossed out the door without the room to explore that hunch (because usually it's a hunch.) Startups, academia, and 21st century corporate R&D are rarely set up incentive-wise to enable to run that hunch to ground. You *do* see this happen with software startups because it is *much* easier to do targeted piddling around with software. Arguably this could be one reason that we've seen less stagnation in software.

In contrast to many modern corporate R&D labs, Google Brain feels like it does the same thing. And perhaps that's why it feels like a much more 'healthy' corporate R&D organization. In fact a lot of the pure AI work feels a little bit like this targeted piddling around. Except that they actually don't take it the one step farther to be an experimental product.

'The Valley of Death' is a concept that pops up repeatedly in discussions about technology development.²⁶ It's nebulous and overloaded, but it always refers to a situation where there is secretly much more/different work than you'd expect between a development stage that feels 'complete' and its natural next step; often this is between a proof-of-concept and a prototype or between a prototype and a manufactured process. I don't find the concept particularly useful, but the thought that "targeted piddling around might be the equivalent of 'hanging out' in the valley of death" feels generative.

Industrial labs enabled smooth project rampups with high ceilings

A common pattern in golden age industrial labs was an individual or small team piddling around with something for a while, then roping in a few people part time, and if it shows promise going from there. Less than a dozen people were working on transistors at Bell Labs from 1939 to 1947²⁷, but within two years,

²⁶ For a good short article on the Valley of Death and its relationship to Technology Readiness Levels, see [Technology Readiness and the Valley of Death](#).

²⁷ Though this time period is probably longer than it would have been because it was interrupted by a world war.

many dozens of people across multiple teams were figuring out how to manufacture them at scale and where they might fit within AT&T's system. Industrial labs were unique in their ability to enable research to start small and then ramp; smoothly transitioning along the technology readiness curve from promising experiment, to prototype, to end-user-quality product.

These smooth rampups stand in contrast to both academic labs and startups. Projects can absolutely start small in academic labs but they will hit a ceiling quickly both because of grant dynamics and publishing pressures.²⁸ Individual grants rarely break the single-digit-million dollar mark, which can only support a small team. Grants are hard to combine to fund a single project because most grants want to support a discrete piece of work. Notably, these grant dynamics mean that national labs and other non-university organizations that depend on grants are under the same constraints. If grants are the input to academic labs, publications are the output. Academic careers are built on citations, but the larger teams required by scaled-up projects either 'dilute' citations or the work isn't even novel enough to publish.

Small businesses don't have the same ceiling as academic labs, but it is hard for them to scale smoothly unless the work produces near-term profit.

Unfortunately, some of the most interesting work precedes profits by a significant chunk of time (even if it will be profitable eventually.) So in order to pursue that work, a small business would need to transform itself into a high growth startup. Startups have the opposite problem of academia: they have a hard *lower* bound on scale. They're expected to grow quickly — if a startup is not constantly hiring it's a red flag.

Ideally, the combination of academic labs and startups should enable this smooth ramping, allowing a project to pupate from academia to a startup when it's mature enough. However, there is often a gap between where you can get in an academic lab versus what you need to start a startup. This is where the Valley of

²⁸ I will only touch on academia's constraints lightly here - only enough to draw the contrast to industrial labs. More on this later!

Death rears its head²⁹ again!

It's informative to compare Google Brain and (the organization-formerly-known-as-Google-)X in the context of smoothly ramping projects. Google Brain has a ton of different things being tinkered on at once, with projects of all different scales. By contrast, X famously prides itself on 'killing ideas quickly.'³⁰ Hard early gating processes prevent the sort of long-burn early work that led to innovations that range from transistors to nylon; hard early gating processes are a good move for startups but perhaps squanders the advantages of Corporate R&D.

A quick research diversion! It would be fascinating to plot a histogram of project sizes at different organizations. My hypothesis is that the organizations that seem to be doing the best solutions R&D would have a nice smooth decay - lots of little projects and a few big projects and many in between. I suspect that many organizations would have a hole in the middle. Maybe that's the Valley of Death appearing as a statistical phenomena.

Prototyping needs manufacturing in the room

'Manufacturing' is code for the folks who will produce the thing at scale.

It's easy to think that if the first version of an invention works consistently, you can just turn around and make a bunch of them. However, the *way* something is made has a huge influence on its cost. A piece of metal that a skilled craftsman shaped by hand is *much* more expensive than that same piece of metal cast in a mold. Sometimes it's straightforward to turn the former into the latter. However, sometimes you need to redesign the thing almost from scratch.

The difference between these two situations is often not obvious to someone without manufacturing experience. There are many non-obvious design choices

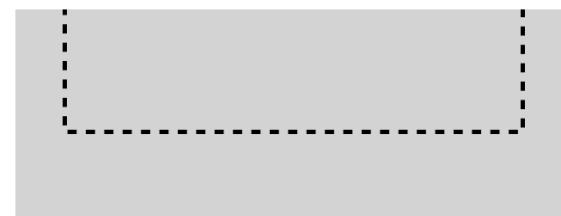
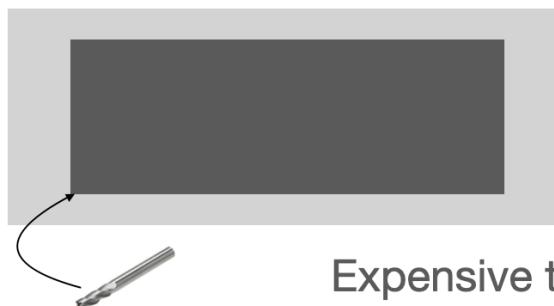
²⁹ It's worth spending more time thinking about the 'dimensions' of the valley of death - because it's not a matter of number of people as the "ramp" language here might imply.

³⁰ See eg. Google X and the Science of Radical Creativity

that make scaling easier or harder. For example, perfect right-angle corners in a rectangular cut-out are completely free (and the default!) in a CAD model, but are *extremely* hard to machine precisely. Often you don't actually need the interior corners of the cut-out to be perfect right angles, but incorporating that tolerance requires shifting other pieces of the design to accommodate it without changing functionality. These design shifts are fairly straightforward if you make them while prototyping. However, once the entire system is roughly complete, manufacturing-focused-design shifts can have cascading effects that require redesigning many other components as well. This is often what's going on when there's a weirdly long gap between a company triumphantly showing off a prototype or concept version of a product and actually putting something in customers' hands.



Cheap to manufacture



Expensive to manufacture!

Round tools don't like
making square corners
:(

Not only are the impediments to scaling non-obvious, they're often tacit knowledge. Instead of a legible rule of thumb like "avoid precise angles at the bottoms of cut-outs" as in the example above, someone with a lot of experience will take one look at a design and say "mm nope that's a bad idea." It's not like they could have given you an explicit list of do's and don'ts — most tacit

knowledge is illegible and even if you could tease it out, the list would be infinitely long.

The importance of having manufacturing in the room applies in software as well as hardware. Different implementations of the same algorithm can parallelize fine or completely break, a functional piece of code could leave giant security holes, etc.

Healthy industrial labs require a trifecta of conditions

A pattern starts to emerge when you ask “how do ‘golden-age’ industrial labs differ from corporate R&D today?” Corporate R&D no longer seems to have the impressive output or cache that it once did.

The golden-age industrial labs shared a trifecta of conditions:

1. They were run by a monopoly.
2. They were working on a clearly high potential technology.
3. That technology addressed one or more existential threats to the company.³¹

Digging in:

1. They were run by a monopoly.

More specifically, they were run by companies that were extracting monopoly rents on some high-margin product. Xerox was the only game in town for copying machines (which, believe it or not, were as essential to many businesses

³¹ Condition #3 was originally “1+2 are insufficient if the technology the lab works on isn’t tied into the company’s core business.” However, this isn’t quite right. There are situations where technology work on the core product can have nothing to do with existential risks to the business (for example Salesforce.) Similarly there are situations where technology work has little to do with the core product but addresses an existential threat (for example flashy research at Bell Labs keeping regulators from breaking up the monopoly.)

as Excel is today), Dupont controlled well-known materials like Teflon or Lycra, Corning was Pyrex and Silicone, and of course AT&T controlled the telephone networks.

2. They were working on a clearly high potential technology.

The technologies the labs worked on generated strong conviction that if researchers could actually pull it off technically, the parent company would be able to sell it. “If you can replace a vacuum tube with something that uses 1000x power, of course people would use it.” “If you can create glass that won’t shatter when you heat it up and cool it down, of course people will use it.”

What’s not clear is why so many fewer technologies seem to fall under that umbrella now. It could be that we picked the low hanging fruit in atom land. It could be that people are just more pessimistic about what we can build. It could be that timelines have shrunk, and there is no high-potential technology on those timelines. It could be that people think in terms of products instead of technology. Or it could be that companies have scoped down so many fewer technologies fall into “things the parent company can sell.”

3. That technology addressed one or more existential threats to the company.

Conditions 1 and 2 are insufficient if the technology the lab works on doesn’t address existential threats to the company. This third condition is often ignored but perhaps most important! In order for an industrial lab to have the massive impact we associate with the greats, it needs active help from its parent organization to bridge the huge gap between an invention and widespread impact. Someone needs to put in active effort to diffuse technology into the world and if the company doesn’t feel intense pressure to do the diffusion work, the technology will languish.

For companies, leveraging core organizational capabilities for a new high-margin product is one of the most straightforward and common ways that new technology can address existential threats. Products become commoditized over time and companies need to keep growing to keep their valuation high. Thus,

new technology can address the existential threat of low margins and flagging growth. However! If that technology isn't in line with core capabilities, taking advantage of it would require significant organizational changes which pose a different (at least perceived) existential threat. So only technologies aligned with an organization's core capabilities address existential threats. This core-business-alignment may also have contributed to industrial labs' impact fulness beyond just enabling them to exist because a technology that can take advantage of existing manufacturing and distribution channels is just more likely to be impactful.

The different dynamics of Xerox PARC's work on the laser printer and the personal printer are illustrative. While PARC is legendary for the impact of its personal computing work, structurally most of that impact shouldn't have happened. PARC's personal computing work *wasn't* tied to Xerox's core business. Different technologies require different organizations and Xerox wasn't set up to scale up and build a *business* around personal computing. The only reason it was impactful was because of a crazy set of contingencies - Steve Jobs' 'raiding party' and Bill Gates quickly following suit.

On the other hand, the laser printer was directly in line with Xerox's core business. Xerox already sold printers/copiers and the laser printer, while a revolutionary new technology, was effectively a better version of those existing products. Xerox could leverage its management structure, manufacturing apparatus, and sales channels to diffuse laser printers with relatively few alterations.

Through the lens of aligning with core capability, Bell Labs practically cheated. AT&T's core capability was the full stack of "communications" so everything from chemistry that kept telephone poles from rotting to information theory that enabled more calls to fit into each wire were actually tied to the company's core business. Transistors had the potential to integrate directly into AT&T's system and enable faster, cheaper service. Faster, cheaper service was an existential priority for AT&T because if they didn't continue to improve, the US government

would have an excuse to break them up. .

I suspect that the existential threat criterion is a big factor in the difference between Google Brain and the-organization-once-known-as-Google-X. Better machine learning technology directly ties into Google Cloud services, search, gmail, etc. While drone delivery and autonomous cars can potentially be big businesses, they are basically orthogonal to Google's core business.

These criteria provide a springboard to briefly explore some abstract but important concepts that seem to apply beyond industrial labs to innovation organizations more broadly.

Innovation organizations need a money factory

Innovation organizations cannot depend on their outputs for free cash flow. A core part of what makes them innovation orgs in the first place is that they create things riddled with Knightian Uncertainty³²that aren't necessarily products. Therefore, they need a consistent external funding source, a *money factory*³³if you will. These external sources can be anything from repeated equity-based investments to a budget from a parent org to an endowment, grants, contracts, or something else.

Don't all organizations need a funding source? Well, yes. For your standard 'vanilla business' that funding source is just the revenue from producing a good or service and then selling it. Their cash flow is directly coupled to their output. Many organizations use financial tools to pull those cash flows from the future to the present (with some discount.) These tools introduce different levels of decoupling from outputs — from still-tightly-coupled short-term loans to

³² I'm using Knightian Uncertainty in its broader sense — situations where there is not only uncertainty but you don't even know what the probability distribution is or even what axis to measure it on.

See Risk, Uncertainty, and Profit.

³³ Some people use the term money fountain - I think this makes it sound a bit too magical. Money factories are mystical enough.

loosely-coupled equity investments. The latter case is so decoupled from output that equity investments are a common go-to-money factory.

There are two major differences between an innovation org and other organizations that forces them to decouple funding from output: They have a significantly longer cycle time and much more uncertainty about their output. Additionally, that output isn't necessarily a product that can be sold directly — it might be a prototype, a process improvement, or simply something that is hard to monetize without hamstringing its impact.

Historically, it takes a long time for actually new ideas to become a valuable thing out in the world.³⁴ This development time means that even if a project is going to become incredibly valuable, the organization creating it is going to be illiquid for a long time. However, the difference between innovation orgs and other businesses is not just about timescales, but also about uncertainty. Large energy and infrastructure projects can take years to start making money but the amount of money they need to get there is fairly predictable³⁵, so they can theoretically get the project done with one lump of money and don't need a *consistent* source of cash. In addition to uncertainty about how long it will take to pay off, new types of work differ from other long term projects because of uncertainty about whether it even *will* generate a return or whether seeking that return is even a good idea if their goal isn't pure profit.

The more an organization looks like past organizations, the more accurately one can predict future performance based on a set of leading indicators. By contrast, it's not clear at all what success metrics should be when you're creating something new. Unlike other long-timescale organizations, an innovation organization might *never* converge on a set of metrics! It's almost tautological

³⁴ Figuring out how to develop technology faster is so impactful because it breaks this initial assumption and weakens the dependence on a money factory. At the same time, technology R+D+Diffusion timescales seem weirdly robust, so the burden of proof is on anyone who claims to have sped up the process.

³⁵ There is, of course, an entire literature about projects like this going drastically over budget and timeline.

that if they are consistently trying to create appreciably different new things then there will consistently be new ways to evaluate those things.

Despite (and in part because of) their inherent uncertainty, innovation orgs need their funding to be *stable* in addition to consistent. The stability may be just as important as the amount. People's incentives go haywire if they have money for now but don't know that it will continue. It might seem that a solution would be to fund individual projects instead of an organization. Unfortunately, uncertain timescales mean that it's extremely hard to do one-shot funding for any given project.³⁶ Innovation organizations are divas: they want consistent cash flows for inconsistent results. Through this lens, it makes sense why many rational people are hesitant to fund them in the way that they need.

An innovation org's money factory needs to deal not just with uncertain timescales and illegible leading indicators but the fact that a chunk of the money is inevitably going to be "wasted." History shows that people are shockingly bad at predicting which research projects are going to be wildly valuable (either in the cash and impact sense) and which are going to be duds. Complicating the matter is the fact that expectations of success or failure can feed back and either hurt *or* help the project.³⁷

Just to recap: Innovation organizations need stable cash flows to support long term projects with significant Knightian Uncertainty and illegible success metrics. Despite their diversity, these common characteristics make it useful to lump funding sources for innovation orgs together into the idea of a "Money factory."

³⁶ Although better program design could help!

³⁷ It would be fascinating if someone did a study on the full two-by-two: situations where expectations of success helped a project, hurt that project and where expectations of failure did the same.

Innovation orgs need to be aligned with their money factory

Aligned incentives between an innovation org and its money factory is the only way an innovation org can avoid crushing oversight and have the ability to work on long-term projects. Playing the same game is the only way people can align for extremely long timescales so if the innovation org is going to survive long term, it needs to be playing the same game as the money factory. The alignment also needs to be clear to the people who control the money and be on a timescale that's acceptable to them. While deep research into the mysteries of the universe might be in the long-term interest of the United States Government, the timescale that matters to most politicians is their term in office. One of the reasons Bell Labs was able to be such an outlier is that AT&T's government sanctioned monopoly and purview over all things communication meant that they were truly aligned with a large range of research.

The more fund-strapped the money factory is, the tighter the alignment needs to be. When AT&T, Microsoft, or Google are flush with monopoly money they're happy to let people piddle around on whatever they wish. When stocks dip, the first programs to be cut are the ones that have the least plausible ties to the core product. Expensive research needs to address an existential threat eventually at an organizational level to maintain support. Similarly, ARPA became DARPA in 1972 because of the increased scrutiny on military spending both in the government and outside of it.

Alignment requires existential threats

What does it actually mean for an innovation org to be aligned with its money factory? When not deciding what their chaotic neutral bard will do when confronted by the town guard, people often use "Alignment" as a fuzzy suitcase word. It doesn't need to be stuffed to the point of meaningless though — as a useful concept it boils down to the blunt question "is maintaining this

relationship holding off some existential threat?"

It's useful to think about alignment in terms of James Carse's concept of "finite games" and "infinite games."³⁸ Each person and organization is playing different 'games' — whether it's exploring the world or maximizing profits. An existential threat is something that potentially brings your game to a crashing halt. Bluntly, the only way for two entities playing different games to have the same goal is if one or both of the games would end if that shared goal isn't met. That is, the goal addresses an existential threat. In the (grossly oversimplified) case of industrial labs and corporations, corporations are often playing the "maximize profit" game. The labs are often playing the "create sweet technology" game. In reality, the two can only be aligned if the labs' intermediate goals either directly help the corporation maximize profit or enable the corporation to keep playing their game (by deflecting anti-trust litigation, for example.)

Of course, 'existential threat' is an extreme term that means different things for different entities. You could think of an existential threat as an event or series of events that could end something that you really do not want to end. This 'threatened thing' can vary wildly in different contexts. For companies it could be an important revenue stream, or the business itself. For an individual it could be your life or just the trajectory of your career. Clearly, what is existential is relative and there's a continuum of importance in the threats.

While it's abstract, I find this concept useful because it allows you to roughly analyze 'how aligned' two entities are and have blunt conversations about it.

³⁸ See Finite and Infinite Games

Expensive research needs to address an existential threat eventually at an organizational level to maintain the support it needs to be effective

If alignment requires existential threats and innovation orgs need to be aligned with their money factories, it follows that that research efforts need to address an existential risk to maintain support. I want to dig into what that means practically.

(A brief aside) I'm going to focus on the realities of how innovation organizations continue to get money in the door. It's easy for the discussion to go down the track of "Painkillers vs vitamins! If I'd asked people what they wanted, they would have said 'a faster horse.' Make things people want! Is physics worthwhile if it never becomes a product? etc." These are questions of "what should a research organization do?" which is important on its own right. However, the less-discussed question is "what should a research organization do *to stay alive?*"

I wish this section's header could be "research organizations³⁹ need to address existential threats to maintain support." The statement is both simple and bold; there are plenty of examples of research orgs that lost support because they didn't really address an existential threat (Dynamicland and BP Venture Research come to mind.) However it's also not true: there is a large class of research organizations that maintain support but continue to limp along ineffectively (Bell Labs still exists!) So perhaps "effective research organizations need to address existential threats" is more accurate. This would explain the difference between ARPA-E and DARPA and why Bell Labs declined once it no longer staved off regulators, among other examples.

However! There are many examples of effective research organizations that did

³⁹ I suspect these arguments apply to innovation organizations in general but I want to narrow the scope here a bit to just research organizations

not address existential threats. In fact, most great scientists were not actually out to address existential threats — Galileo, Newton, Rutherford, Einstein, etc. They just managed to cobble together enough money from patrons or side hustles to keep going. Patreon-sponsored contemporaries are similar — I give some money in support of a few researchers but they aren't addressing any existential threat for me. The notable pattern is that these examples are all individuals or small groups. Aha! It suggests that perhaps there is a nebulous threshold below which effective research can work off of 'throw away' money and above which people start looking at money spent on research as 'buying' something — we could call this point "expensive."

Another potential Achilles heel in this idea is the fact that often the work that convinces us outsiders that a research organization is effective is often some of the *least* existential-problem-addressing work that the organization does: transistors at Bell Labs, interactive computing at DARPA, etc. However, at the same time that it was inventing the transistor, Bell Labs was discovering better wire sheathings that saved AT&T billions of dollars and DARPA was working out ways to detect nuclear explosions anywhere in the world at the same time that Licklider was seeding interactive computing groups across the US. This tension is worth noting because it's important to not lash any single project too tightly to the importance of addressing existential threats at an organizational level.

So, the correct statement is perhaps that *expensive* research needs to address an existential threat *eventually at an organizational level* to maintain support. This quite the mashup of nebulous words isn't it? It's important enough that it's worthwhile to dig into each piece.

What does it mean for research to be 'expensive'?

The line delineating 'expensive' is relative and slippery, but there's definitely a line. In large part its psychological. Buying something expensive causes you to pause and consider. You care much more about its outcome. While your wealth level absolutely factors into what is expensive and not expensive, it's nothing like

a 1:1 relationship. Someone who agonizes when cabbage prices go from \$0.69/lb to \$0.89/lb will buy a \$500 phone without thinking about it. This same phenomenon happens with corporations, governments, investors, and philanthropists as well.

What does ‘eventually’ entail?

Most organizations have a ‘grace period’ to piddle around. However, like the border between expensive and not-expensive the length of this grace period is extremely fuzzy and liable to change at a moments notice. Usually when it’s confusing why a corporation or investor would be supporting work that is so far from a product, it doesn’t mean you’re missing something, it means the hammer is about to fall. It’s almost indescribable, but there is a distinct sense you get from a startup or new research lab that suddenly runs into the end of its grace period — it goes from feeling like Willy Wonka’s Chocolate Factory to ACME Chocolate Improvement Inc.

“Eventually” also raises a question about the frequency of hits that an organization needs to maintain, which seems to be a function of how big those hits are and how existential the threats they address actually are. DARPA has a 5-10 percent program success rate, which I suspect it can get away with because military superiority is *very important* to the US government, while most charities need to show progress every semi-annual fundraising season.

Why do existential threats need to be addressed at an organizational level?

The *organization* needs to address an existential threat to its money factory,⁴⁰ rather than any individual project. Effective research organizations seem to build a portfolio of projects that address existential threats at a sufficient rate to maintain the perception that they’re addressing existential threats.

⁴⁰ Or at least be perceived to be addressing existential threats - there are many organizations that successfully maintain funding based purely on great marketing. It’s worth paying attention to why this marketing works but I’m going to otherwise ignore it because I am bad at pulling off false pretenses and would prefer if the world had fewer of them.

This portfolio approach is clutch for several reasons. There is uncertainty around any given project as to whether *it* will address existential threats and often you can't honestly answer that question a-priori without doing some work or strangling the project in the cradle.⁴¹ Additionally, the projects that directly address existential threats for funders and those that are most interesting or globally impactful are often disjoint. AT&T didn't overwhelmingly benefit from the transistor or cell phone nor did Xerox benefit from the GUI. In the same way that winners pay for duds in a VC portfolio, aligned work can cover for misaligned work if the alignment is considered at an organizational instead of a project level.

If the organization produces aligned hits at a sufficient rate, it ideally can create a trusted hierarchy where the trust (and money) is flowing *down* the power ladder. Congress trusts the DARPA director who trusts the deputy director who trusts the PM. The PM doesn't need to get congress' permission to start a program, just the permission of the deputy director.⁴² This is why opacity is important to DARPA's outlier success — it can work on crazy ideas that would never get a-priori funding from congress, but it needs to continually renew the trust to maintain that opacity by delivering as an organization.

Another way this manifests is through contract research organizations that fund themselves through contracts or grants, but use that money to fund internal research.

Industrial Labs no longer fill the niche they did in the early-mid twentieth century

In the past, industrial labs filled a particular niche in the innovation ecosystem.

⁴¹ The book Loonshots introduces the concepts of “warty babies” and “the three deaths” which are incisive mental models of the phenomena that lead to project-cradle-death.

⁴² This is very true of 1960’s ARPA but less so of DARPA. Over time, congress has exerted more direct oversight over DARPA programs. That being said, there’s still less overhead than other government research.

There's a strong sense that as of 2021, corporate R&D organizations⁴³ no longer fill that niche. However, these organizations (including Xerox [PARC](#) and [Bell Labs](#)) still exist, so we need a strong argument that they no longer fill the Solutions R&D niche.

I'll proceed by both looking at outputs and inputs. Industrial labs are no longer producing the real-world outcomes they once did. However, that argument has a causality problem — it could be that labs are actually filling the same niche and a confounder is causing decreased output. This confounder could be the same thing driving stagnation in general. For example, if we really have picked a finite number of low hanging fruit and Robert Gordon is right that the explosion of growth between the late 19th and mid 20th century was a one-off event. I want to argue that instead, the conditions that enabled industrial labs in particular to do great work no longer exist.

There are many exceptions! AI research in particular is a glaring exception to the decline of industrial labs but there are many potentially amazing corporate R&D projects happening around the world.⁴⁴ One might argue that the exceptions *are* the rule — that there are just fewer technologies that would benefit from healthy industrial Labs (remember, the Low Hanging fruit confounder.) It's impossible to prove definitively, but I want to argue that the niche industrial Labs once occupied still needs to be filled.

Going through the characteristics of the niche we discussed previously:

- **Most industrial labs no longer work on general-purpose technology.** The trend, as embodied by the 'founding memo' Nathan Myhrvold wrote about Microsoft Research.⁴⁵ In it, he explicitly focused Microsoft Research on advanced product development and not cutting edge research. This trend is also seen at the-organization-formerly-known-as-Google X where they pride themselves on killing ideas quickly and then focusing on a specific product. The trend towards

⁴³ It's unclear whether it they should even be called labs!

⁴⁴ Lockheed Martin is supposedly working on a portable fusion reactor,

⁴⁵ The memo is quite good and available [here](#).

product development pushes corporate R&D more towards the same niche filled by startups.

- **Industrial labs no longer enable smooth transitions from crazy proofs-of-concept to manufactured products.** Either the labs are so incredibly separate from the main organization that they have no advantage over a startup or are so tightly integrated that anything new is quashed. In the first case, there is no systematic way for projects to move from the lab to the main organization. Instead, they are forced to go through the same process of ‘finding a champion’ and fighting a bespoke uphill that any product going to market faces. These barriers create the same valley of death that the academic project spinning out into a startup path faces. In the second case, labs are so joined at the hip to the main org that they can’t work on anything besides incremental improvements.⁴⁶

- Echoing the previous point, **industrial labs often no longer work on potentially game-changing technologies with tight feedback loops.** They’re either doing “show-off” work that leads to nifty demos and papers or grungy incremental work. To posit a reason why: technology improvement is no longer an existential imperative for most corporations. Today it’s much easier for corporations to acquire new technology through M&A because of relaxed antitrust and a vibrant startup ecosystem. Modern corporations also face very different pressures from AT&T in the mid 20th century. AT&T benefitted significantly from creating better technology both because it was the key blocker to their expansion and the anti-trust hammer was always waiting to drop if it didn’t appear to be benefitting the public.⁴⁷

- **Industrial labs no longer provide a first-class alternative to academia.**

With the exception of the big AI labs and maybe IBM research on superconductors, there are few areas where working at an industrial lab is as

⁴⁶ For more on the first case, see the section in Loonshots about “Manage the transfer, not the technology” and for the second, see “Separate soldiers and scientists” (Have I convinced you to read it yet?)

⁴⁷ See Killing the Golden Goose? The changing nature of corporate research, 1980-2007 for an in-depth exploration of this trend.

academically prestigious as being a professor. Like the shift away from game changing technologies, I attribute the declining prestige to the fact that cutting-edge science and technology are no longer an existential concern for companies. The absence of an existential threat gives them little reason to hire and keep the best and most ambitious researchers, creating a downward spiral where there are no A players and the labs become susceptible to the fact that A players hire A players and B players hire C players.⁴⁸ Additionally, the incentive shift to focus more on product development has created a cultural shift away from participating in the scientific enterprise.⁴⁹

- **Industrial Labs no longer enable continuous work on 5+ year timescales.**

There is a lot of literature on corporate short-termism that I won't dig into. In short, most companies that have the resources to do research on a 5+ year timeline are disincentivized from speculative 5-10 year projects through a combination of focusing on quarterly earnings, competition and shareholders pushing companies to stick to their knitting, and the availability of M&A opportunities. Companies fail to support long-term research not just in the obvious sense that the projects get cancelled or never started in the first place. The more insidious and common failure is that there's no level of planning for how promising projects eventually become part of the core company so even if they're not cancelled they putter along forever.

- In most cases, **industrial labs no longer probe deeply into natural phenomena and try to exploit them.** Industrial Labs just don't do as much *science*. This isn't to say that industrial labs ever pursued scientific discovery for its own sake, but in the past they perceived a direct line between scientific discoveries and the health of the company. I suspect that this perception is accurate in most cases! The move away from Pasteur's quadrant could simply be because we've exhausted exploitable phenomena and everything left is either too

⁴⁸ Before you write me an angry email, I am not saying that all teams at all industrial labs are now C players, but it is undeniable that most people *on the academic track* don't view industrial labs as a first-class alternative to universities.

⁴⁹ See The decline of science in corporate R&D for a more in-depth look.

big, too small, or too obscure to make a difference at human time-and-size scales. Alternately, the move could be chalked up to a change in the nature of the companies and their products to be more specialized and less dependent on exploiting new phenomena.

AI research is an exception to the decline of industrial labs

AI research bucks the trend of the decline of industrial Labs as a place where globally important, cutting-edge work is done. If you squint, Organizations like Facebook AI Research , Google Brain, DeepMind, and possibly OpenAI⁵⁰ bear strong resemblance to Bell Labs, Xerox PARC, and other legendary industrial labs in their heydays. It's worth understanding why this resemblance exists in these labs but not elsewhere to understand whether 'creating the new Bell Labs' is a reasonable goal in other disciplines.

AI research can require massive resources — in this case it's the thousands of dollars of compute for training models and creating massive datasets to train on. These resource requirements mean that there is exploratory research that people just can't do with the resources available to most labs at universities. As a result, industrial AI labs are a first-class alternative to academia with titans of the field like Yann LeCun and Samy Bengio joining FAIR and Google Brain respectively. A similar effect pulled professors away from universities and into industrial Labs in the first half of the 20th century.

At the same time, corporations running AI labs (reasonably) expect the research to create value commiserate to the costs of this research. Machine learning can directly improve the core product lines of all the companies listed above, in the same way that Bell Labs' work directly improved AT&T's 'System'. Additionally, AI promises massive value over long but not infinite timescales. OpenAI's business model is implicitly based on the assumption that they will create

⁵⁰ At this point OpenAI vaguely resembles the Bell Labs equivalent for Microsoft.

literally infinite value in the not unforeseeable future.

This alignment between their research and money factories allows modern AI labs to do highly-regarded work and collaborate closely with academia without it seeming like a waste of money. Bell Labs provided a home for work that led to nine Nobel prizes, often in collaboration with universities. Today, AI conferences are dominated by work from corporate AI labs, not just in quantity but quality as well.

Like the electronics work at Bell Labs, AI research benefits from having multiple disciplines in the same place with the people who will eventually produce it. When you're creating "AI for <X>", it's generally helpful to have someone who is an expert in <X> around. Training and implementing ML algorithms at scale requires a different skillset than creating and prototyping the algorithms. Corporate AI labs have the budget to hire research engineers and also make it easy for researchers to talk to some of the best infrastructure and production engineers in the world. This dynamic roughly parallels the close ties between Bell Labs and Western Electric and is effectively the software version of "Prototyping needs manufacturing in the room."

Perhaps the most uncomfortable parallel is between AT&T's status as a high-margin, cash-printing monopoly and Google or Facebook's similar situation⁵¹. AI labs provide more evidence that successful industrial labs can only exist in domains where there are large corporations with monopoly-like profits. The decline of industrial labs in chemistry and physics-related domains may have been caused by commoditization of those products. GE, Dupont, Kodak, and others' share prices suggest that they at least are no longer perceived as monopolies. However, monopoly profits -> high quality corporate research isn't the entire story because you don't see a lot of high-quality non-product research coming out of Boeing and arguably Amazon⁵². Perhaps the missing difference is

⁵¹ Yes, Google and Facebook are not technically monopolies in the full AT&T or Standard Oil sense, but the flavor of their profit margin and size is similar.

⁵² Maybe it's a Seattle thing?

whether or not the corporation perceives that it will benefit from significantly better technology. AT&T benefitted significantly from creating better technology because they had a pact with the government that as long as they kept making the system better and did public-benefitting research they wouldn't be broken up. I wonder what the counterfactual is — if the government left them a monopoly without the pact. While Google and Facebook have no (publicly known) agreement with the government, arguably Google's ad-fueled profit does depend on them continuing to have the best search engine — people have no loyalty to google beyond the quality of their searches. Similarly one could make an argument that advertisers will abandon Facebook as soon as they have worse ad targeting or the kids move to another platform like TikTok or Snap.

Microsoft is an interesting beast to examine through this lens. It's unconfirmed, but Microsoft Research (MSR) may have been started as a public offering in the same way Bell Labs was. Regardless, it both explicitly set out to do more product-focused work than PARC and Microsoft's cash cows were Windows and the Office Suite, both of which didn't benefit much from MSR work. MSR was subsequently pretty neglected and used for stunts like China expansion. However, Microsoft's core business has begun to shift towards cloud services, which *do* benefit from AI research and it's becoming clear with things like GPT-3 that AI research can augment Office Suite products that are now actually under threat. This shift coincides with a partnership with OpenAI⁵³ that looks a bit like the relationship between AT&T and Bell Labs if you squint. Amazon remains a challenge to this just-so story because it also has a massive cloud business but is not (publicly) doing as much speculative AI research.

This narrative suggests an uncomfortable truth. Instead of declining because of "corporate short-termism," speculative corporate R&D might depend on companies with monopoly-like profits perceiving their fate being intimately tied to a high-potential technology. This trifecta (monopoly profits + clearly high

⁵³ See OpenAI forms exclusive computing partnership with Microsoft to build new Azure AI supercomputing technologies

potential technology + that technology being tied to the company's core business) can't be created by policy or even a cultural shift. The trifecta is the most straightforward answer to the question "why hasn't Bell Labs been replicated?" and suggests that attempts to "create a new Bell Labs" are doomed to failure if they try to follow the playbook too closely. At the same time, I strongly suspect there are incredibly valuable technologies that don't clearly have high potential to solve an existential threat for a monopoly rentee but do need the environment and resources that were once provided by healthy industrial labs to realize that value.

The niche industrial Labs once occupied still needs to be filled

Even if it's true that in the past, industrial labs filled a particular niche in the innovation ecosystem and that industrial labs no longer fill the niche they did in the early-mid twentieth century, it doesn't necessarily follow that the niche still exists or that it would be good⁵⁴ idea to try to fill it. It's possible that we've pushed human capability to a point that has rendered the solutions R&D niche irrelevant. If we have truly unlocked every phenomenon it's possible to exploit, even Bell Labs in its full power would not have impressive output. However, I suspect this isn't the case.

It's hard to positively⁵⁵ answer the question "is the niche once filled by industrial labs still valuable?" Part of the nature of solutions R&D is that valuable work is often not obvious. If we could point to an area of the adjacent possible⁵⁶ that is

⁵⁴ I'm not going to dig into the philosophy of "what makes something good?" I am an unabashed explorationist, but I'm pretty sure you could make a utilitarian argument as well.

⁵⁵ Positive in the sense of "positivist" not in the sense of "Yeah! Science!"

⁵⁶ The "adjacent possible" (grossly simplified) the possible states that a system can enter, given its current state. In the case of science and technology, it's the set of things we could *possibly* invent and discover given what we've already invented or discovered. The concept was originally introduced by Stuart Kauffman in his book *Investigations*

uncontroversially valuable, it arguably is no longer even solutions R&D. So it's hard to point to specific, clearly valuable solutions R&D areas that are not being tackled. Instead, I'm going to approach the question with the double-negative argument that we should *not* believe that the reason the niche has been vacated is because it is not valuable. I realize that's a bit convoluted. It's like a negative-space painting where I cannot portray the subjects directly, but instead suggest their existence by filling in everything else.



First, let's take seriously the idea that the solutions R&D niche isn't valuable we've picked all the low hanging fruit that were once harvested by industrial labs. It has indeed been many years since we have unlocked a new physical phenomenon for exploitation in the same way we tapped the electron and steam before it.⁵⁷ It could be that we've exhausted exploitable phenomena and everything left is either too big, too small, or too obscure to make a difference at human time-and-size scales. If this is the case, the enterprise of discovery has become decoupled from the enterprise of invention and while there are big discovery projects to be done, they will be purely for curiosity's sake. If this is the case, all the possible new inventions shouldn't require a lot of piddling around in disciplines that border on science. This argument is plausible. After all, companies with science-based products still have labs: 3M, Corning, Kodak, and

⁵⁷ "The Cell! DNA!" You may shout. It's hard to argue that we have benefitted from manipulating DNA as much as we did from manipulating electrons. Biology certainly has that *potential*, but so did nuclear physics ...

others. One would expect that if there were fruits to be picked, these labs would be picking them. Instead it looks like they're hanging out at the top of their respective S-Curves and not producing new ones. Dupont gutting its once-legendary R&D department⁵⁸ could be seen as a concession to this new state of the world.

However, there is also evidence a world where industrial labs no longer fill the niche they once did was created by forces that were driven by shifting economic and cultural factors more than a depleted adjacent possible.

[The decline of science in corporate R&D](#) argues that increasing international competition and pressure from shareholders to specialize forced companies to decrease the scope of corporate R&D and shift towards more directly product-related work. The Microsoft Research Memo is evidence for this scope-shift — managers realized that Bell Labs and PARC-style R&D orgs were not actually great for business. Jettisoning industrial labs from a purely financial standpoint also makes sense when there are many perceived M&A opportunities to acquire new technology instead of developing it in-house. The decision is further supported when labs no longer function as monopoly-preserving peace offerings.

Few companies today meet the trifecta of conditions that healthy industrial labs require, and those that do are existentially tied to bit-based technologies more than atom-based technologies. The companies with science-based products that still have labs no longer command massive high-margin profits, forcing more conservative work which doesn't push far enough into the adjacent possible to unlock new high-margin S-curves. It's possible that the conditions for healthy industrial labs that focus on atom-based technologies could occur again but creating those conditions requires so many other contingent things to happen first that it won't happen anytime soon. For example, it's possible to imagine a world where a company has mastered a high-margin application of molecular machines and to keep its edge it needs to push ever farther into the realms of

⁵⁸ See DuPont Will Dissolve Central Research

atomically precise manufacturing. However, even if that company were to start today, it would take a decade before it was at a point where it could create a healthy industrial lab.

What about government labs? We haven't talked about them much but government research is basically corporate R&D for United States Inc. If there were valuable parts of the adjacent possible that industrial labs could explore, shouldn't we expect government labs to be filling the solutions R&D niche? Yes and no. The US government technically meets the conditions for healthy industrial labs — it's a monopoly and benefits from value created in the US. As one would expect, in areas where the America feels an existential threat, National Labs and other government research organizations do actually look like healthy industrial labs. However, there are fewer areas where research can address existential risk for a country than you might expect! If you look at national labs through the lens that "Alignment requires existential threats", everything makes more sense; National labs⁵⁹ excel at work related to nuclear weapons and military capability in general. Perhaps a bit cynically, arguably everything else National Labs do is a nice-to-have but doesn't address an existential threat to the government.

It's impossible to rule out the grim possibility that we have depleted the areas of the adjacent possible that solutions R&D can reach. However, there is strong evidence that even if there was fertile ground, industrial labs (including those of United States Inc.) would not be cultivating it. In this case, it's up to other institutions to step boldly into the void.

⁵⁹ Which, like the DOE itself were originally created for the sole purpose of smoothly transitioning nuclear weapons from proof of concept to manufactured product!

We now have a sense of what the solutions R&D niche looks like and have come to grips with the fact that while industrial labs have abandoned it, there are still potentially many unpicked fruits within. A common view⁶⁰ is that a combination of academic labs and silicon-valley-style startups is sufficient to fill that niche. While academia and startups both wander into the periphery of solutions R&D, its core activities run headlong into their institutional incentives. The work an organization needs to do to occupy the niche is fundamentally too engineering-heavy for academia and too researchy for startups. In this section, I unpack how both academia and startups are constrained out of filling the solutions R&D niche. Startup constraints touch on a broader discussion of the tricky relationship between profit, value capture, and impact of research and innovations. Legal structures are intimately tied to both the question of money and institutional capacity, so I lay some groundwork there as well. It might seem like traditional nonprofits should be able to avoid this entire mess, but that's not the case when dealing with long-term, high-uncertainty, relatively capital-intense projects. As a consequence, we need to go a bit into the unknown and create new institutional structures for doing solutions R&D.

In the latter 20th century and early 21st century, industrial labs yielded more and more of their niche to a combination of academia and newfangled “startups.” You can see this change in data on scientific publications (which have gone through the roof in academia but steadily declined in companies, suggesting that researchy work has consolidated in academia) or the number of mergers and acquisitions (also through the roof, suggesting that large companies have increasingly outsourced early-stage engineering work to startups).⁶¹ A less rigorous but more powerful way to see the change is by ‘cultural pattern matching’ — that the sorts of projects that you would once have expected to find in an industrial lab are now either in an academic lab or a startup. Many people

⁶⁰ This view is sometimes implicit, but often explicit.

⁶¹ See The decline of science in corporate R&D again

implicitly or explicitly believe that between academia and startups⁶² the solutions R&D niche is filled. I once believed that myself, but came to realize that there are many projects that don't fit within the constraints of either academia or startups. "Things that are too engineering-heavy for academia and too researchy for startups" is shorthand for one class of these projects.⁶³ By unpacking that term, I hope convince you that the combination of startups and academia is insufficient to occupy the niche once served by industrial labs.

Academia revolves around scientific inquiry

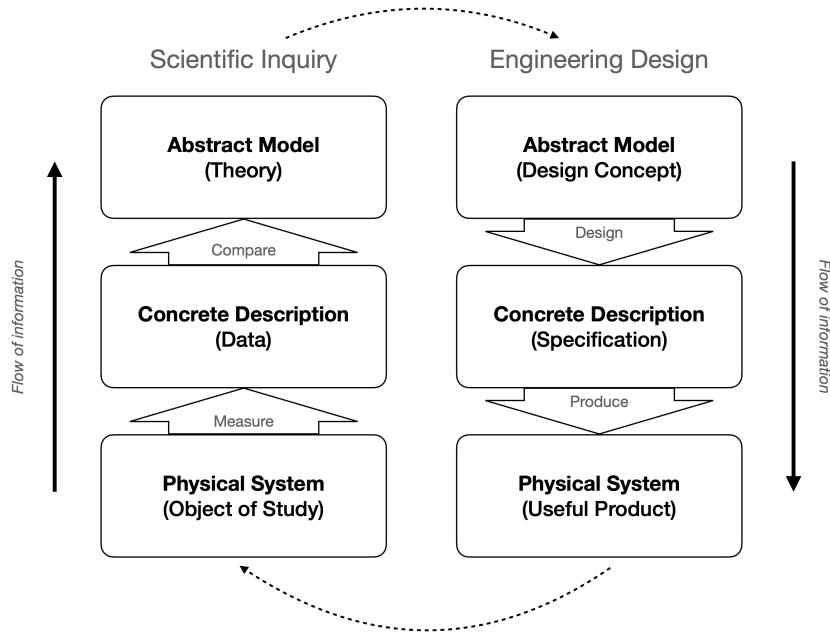
Quick aside: 'Academia' is a nebulous term. It doesn't strictly refer to only universities. Instead, I use academia to cover all the people and institutions who are playing the incentive game where you get points for papers, citations, and general adulation from other people who are playing that same game. In this sense, you can participate in academia partially; there are some professors who are barely part of it beyond their institutional ties; there are people with jobs entirely unrelated to research who nevertheless are driven by academia's incentives.

The academic model is built around scientific inquiry. Academics are rewarded for moving information *up* the ladder of abstraction.⁶⁴ The more general your theory or technique, the more you are praised. Unfortunately, the academic reward system is at odds with the engineering design work you need to create useful products.

⁶² Broadly including institutions in their ecosystems like venture capital, incubators, national research centers

⁶³ There are many others!

⁶⁴ See <http://worrydream.com/LadderOfAbstraction/>



Schematic of scientific inquiry vs. engineering design from *Radical Abundance*.

In [Radical Abundance](#), Eric Drexler presents this incredible chart that captures the contrasting dynamics of scientific inquiry and engineering design. In a nutshell, scientific inquiry is a process where information flows from the concrete to the abstract, ideally resulting in an abstract model — the more general the better; engineering design is a process where information flows from the abstract to the concrete, ideally resulting in a physical system — the more useful the better. Neither of these processes ‘precedes’ the other.⁶⁵

Moving down the ladder of abstraction clashes with academic incentives that have been set up to support scientific inquiry. As a result, academic engineers default to creating proofs-of-concept not as a step towards a working system but for the sole purpose of validating general engineering principles or design concepts. The whole working system is left as an exercise to the reader. Proofs-of-concept, general principles, and designs are important! But they are a long way

⁶⁵ In fact, it would be accurate to draw an arrow from the bottom of engineering design to the bottom of scientific inquiry and do the same from the top of scientific inquiry to the top of engineering design, creating an ouroboros: The mythical snake eating its own tail

from a useful thing out in the world. For context, proof-of-concept is level three of nine on NASA's Technology Readiness Scale⁶⁶. More importantly, a proof-of-concept usually isn't enough for a technology to gain momentum by attracting the money and effort to continue the process of engineering design. Here is where the Valley of Death raises its head again.

Because academic engineering focuses on the top of the ladder of abstraction, feedback loops that can only come from implementation are rare. A design can check all the boxes on a list of requirements (which makes it great for a paper) but completely fail to be a useful system.⁶⁷ My favorite example of this takes us back to the example in Prototyping needs manufacturing in the room: recall how perfect right-angles in cut-outs are completely free (and the default!) in a CAD model, but are *extremely* hard to machine correctly. These weirdly important implementation details are usually tacit knowledge that does not exist at higher levels of abstraction. It gets 'abstracted away.'

Novelty is an important tool for moving up the ladder of abstraction but it can be a wrench in the gears of useful engineering design. When your goal is abstract knowledge, anything that's fully captured by existing understanding is worthless. As a result, academic work is always judged by "what are the new ideas here?" Rather than "does it work?" On the other hand, a reliability and success-maximizing engineering rule of thumb is to use as few new things as possible to get the system to work. This tension means that academic engineers will not work on some ideas or bolt on unnecessary fancy techniques. Welding together a bunch of old ideas to get something to work is great engineering practice, but it doesn't create any general academic insight and thus doesn't lead to papers, promotion, or tenure.

At the end of the day, ideas need to come out of a single mind. The abstract models at the pinnacle of scientific inquiry are effectively just well-justified ideas.

⁶⁶ See <https://benjaminreinhardt.com/trl>

⁶⁷ This can of course happen outside of academia! But in academia it can go forever without running into the cold hard wall that is reality.

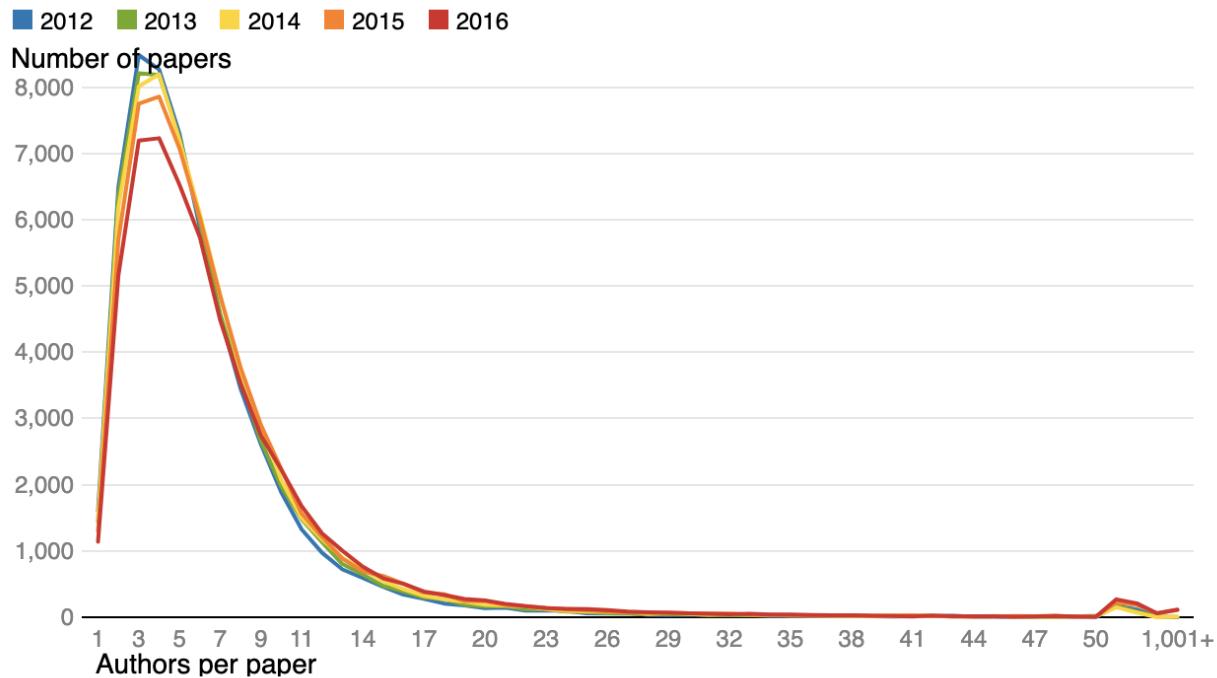
Scientific mythology has further solidified the paradigm of theory leaping forth from a single great mind like Athena from the head of Zeus. Newton, Einstein, Curie... As a result,⁶⁸ academic culture prizes individual recognition, not team output. In academia, citations and (especially first) authorships on papers are the coin of the realm. These status and reward mechanisms push people to make sure each piece of their work is recognized. Many engineering projects require ‘teamwork’ in the sense of putting your head down and doing the unsexy work that needs to be done without remembering who did what. This isn’t to say that academics won’t do unsexy work⁶⁹, but it does need to culminate in something paper-worthy.

There *are* an increasing number of academic papers with large numbers of authors, but there is a roughly constant amount of “esteem currency” per paper so the career capital per author decreases as the number of authors increase.⁷⁰ Contrast the incentives in academia to professional engineering, where reward (both status and monetary) is usually invariant to the number of people on the project. When someone says “I helped build the Falcon 9 Rocket” or “My instrument is on the Perseverance rover” nobody asks “ah, but how many other people also worked on it? Were you first author?” It’s viscerally cool even if you just tested one non-critical part. When “does it work?” is the biggest source of recognition, the incentive is to have as many people working on a project as necessary to make it work.

⁶⁸ And through self-reinforcing cultural evolution.

⁶⁹ See especially: any graduate work that involves culturing cells or working with equipment that is more than thirty years old.

⁷⁰ See *Paper Authorship Goes Hyper*



Only papers included in the 68 journals tracked by the Nature Index are represented.

Source: [Nature Index](#) • Created with Datawrapper

Screennotate Trial

Academic Constraints

In order to build up our pattern-matching, it's useful to explicitly enumerate the constraints that define the nebulous edge of academia's reach. By rights, this list needs its own piece and we touched on several of these points already but they bear repeating.

A lot of engineering work is not novel. Spend enough time in research seminars and you'll inevitably hear many variations on both "our work is novel because ..." and "yes, but it's not really novel is it?" The word 'novel' is used as an idea bludgeon in academia and weeds out many classes of ideas.

A non-exhaustive list of what gets discarded by the novelty filter:

- New approaches to old 'solved' problems.

- Better implementations of an idea that's already been implemented.
- Efficiency improvements along a different metric than the ones the field normally cares about.
- Scaling existing techniques.
- Creating something for a specific application as opposed to a general principle/purpose.
- Further work on things that have already gotten to proofs-of-concept.
- Integrating several already-existing ideas into a single system.

Good engineering does not always get many citations. Even if work is novel enough to be published, low-citation ideas languish in obscurity. It's valid to ask "are there things that are valuable that don't get a lot of citations?" Absolutely. Valuable projects that want to live in an academic lab but wouldn't get many citations include:

- Projects that are intermediate steps to a larger vision without a strong conclusion besides "we did a thing!"
- Engineering work in a theory-heavy field or theory work in an engineering-heavy field. More generally, work in a field that the field doesn't think is valuable.
- Ideas that are far out of mainstream. Even non-malicious people often don't have the bandwidth to dig in and see if an idea from left-field is worthwhile or utter crackpottery. It's usually the latter so the default answer is "no." (see the point about stepping too far outside the mainstream below)

Engineering just asks "does it work?" In academia it's hard to publish ideas that are too outside of the approved way of doing things. Kuhnsian paradigms outline a set of questions and the rules around answering them.⁷¹ Ideas can stray

⁷¹ See The Structure of Scientific Discoveries

too far outside of accepted paradigms when they're asking different questions, or measure success differently than the ways the paradigm prescribes. Boyden et al's work with expansion microscopy⁷² is a good example: instead of asking how "how good can we make a microscope's resolution?" they asked "how big can we expand a sample while maintaining its structure?"⁷³ It's also hard to publish work that doesn't fit cleanly into a disciplinary box⁷⁴. It's too much like discipline X to be published in a journal for discipline Y but it's too much like discipline Y to be published in a journal for discipline X. This situation is of course tricky because many (most?) ideas that go outside of established paradigms are crackpottery.

It's hard to write a paper if you can't even get the money to do the research. The need to align with current funding priorities is particularly pernicious in academia because most grant processes depends on politics and committees. By their nature committees lead to median results, so committee-mediated grants will rarely fund unsexy outlier work. One reason to write papers is to continue playing the game, so if a paper doesn't hint at your continued ability to do work that aligns with funding priorities, there's less incentive to write it.

A project is unlikely to lead to a paper if it's not within scope for a PhD student or untenured professor. Graduate students are the labor force in academic science and engineering. In exchange for cheap labor, grad students expect to produce research within a 4-7 year time frame that can help them move on to the next rung of the academic career ladder.⁷⁵ While a tenured professor might be willing to work on something crazy that will take 10 years and could be

⁷² See Expansion Microscopy

⁷³ But they were published and now have many citations! You say. Yes, after years of struggling to publish and get funding. There is a heavy survivorship bias here because there's rarely anything to point to in situations where something wasn't published.

⁷⁴ I want to use the term "antedisciplinary science" from Sean R. Eddy's article of the same name but relegated it to this footnote to avoid too many new words.

⁷⁵ Despite the fact that there is an increasingly wide gap between the number of graduate students training for PhDs and academic positions for them to fill!

a complete dud, no grad student would rationally work on it and no advisor with a student's best interest at heart would ask them to. The incentive for PhD students to produce high-impact papers for the sake of their careers amplifies all the other constraints; PhD students in science get most of their funding through research grants, so their work is especially susceptible to grant-related constraints.

Projects whose output is structured data or reusable code don't lead themselves to papers. Academic papers are built around the process of scientific inquiry, where collecting data and writing code are only in service to producing an abstract model or theory - the more generalizable the better. There is little reward to producing high quality, reusable datasets or code, despite the value they can create.

A project is unlikely to lead to a paper if it involves a lot of coordination. Academia emphasizes individual agency. This is a good thing, but if a project cannot be effectively modularized it will end up requiring a lot of grungy coordination work that doesn't count as a contribution unless you're in charge of the project. A big reason to go into academia in the first place is to avoid having a boss and because you like doing your own thing, your own way. Obviously there are exceptions like the LHC, but that is a situation where the experimental particle physicists are locked into a paradigm and have no other option.

Coordination aversion hits especially hard projects that don't fit into a particular disciplinary bucket or require integration between components built in different labs. Projects of a particular scale require coordination by their very nature.

Startups revolve around growth

A startup is a company designed to grow fast. Being newly founded does not in itself make a company a startup. Nor is it necessary for a startup to work on

technology, or take venture funding, or have some sort of "exit." The only essential thing is growth. Everything else we associate with startups follows from growth.⁷⁶~Paul Graham

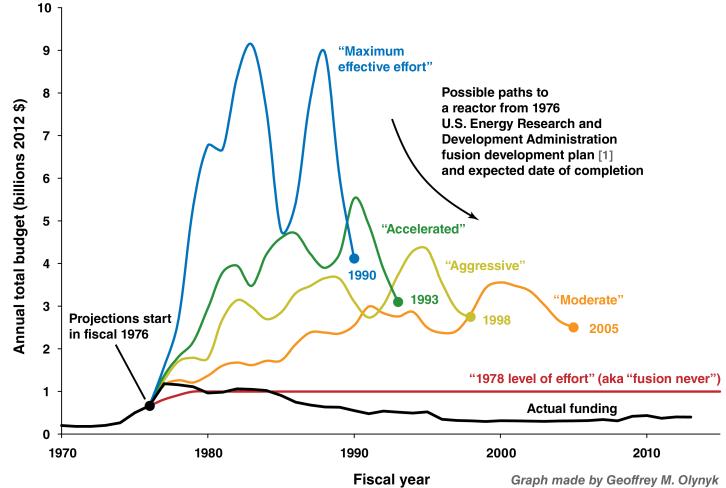
Just as academia revolves around scientific inquiry, startups revolve around growth. In our context — filling the solutions R&D niche, this focus on growth creates several corollaries. Obviously, it *does* mean we're talking about companies that work on technology. And not just software-driven-scalable-business 'tech' but actual technology. A technology startup is extremely unlikely to be able to fund development by selling products from day one. The company then needs to take investment, short of extraordinary circumstances like being run by a multimillionaire or some kind of absurd government contract that pays in advance. The only people who will make that kind of investment are institutional VCs and individual Angel investors. Raising VC money then commits a startup to have some sort of 'exit'. So for our purposes, 'startup' means 'VC-funded startup' along with the consequent constraints.⁷⁷

The coupling between startups and growth raises the reasonable question: "what about low-growth small businesses?" As we noted earlier, the nature of innovation orgs makes them dependent on an external money factory. If a small business is doing self-directed R&D, they can't depend on its results for free cash flow. If the company raises money, it implicitly binds itself to all the startup constraints we will dig into shortly, even if it does not consider itself a high-growth startup. If the company forgoes investment and the founders are not extremely wealthy (eg. SpaceX), the company needs early revenue from somewhere. The common move for bootstrapped companies who need to do more development is to do consulting work to generate revenue that they then plow into development. Unless it's wildly profitable, the consulting approach severely limits the amount of time and money that can be spent on development, possibly dragging it out forever. SBIR grants are another option, but they are relatively small (order \$1m over several years) and have the same effect as

⁷⁶ <http://paulgraham.com/growth.html>

⁷⁷ There will always be exceptions!

consulting with small companies limping from one grant to the next. Some projects need effort above a certain threshold to move forward!



[1] U.S. Energy Research and Development Administration, 1976. "Fusion power by magnetic confinement: Program plan" ERDA report ERDA-76/110. Also published as S.O. Dean (1998), *J. Fus. Energy* 17(4), 263–287, doi:10.1023/A:1021815909065

Startup Constraints

Similarly to the constraints on academia, this list deserves its own piece! This list is meant to enumerate the constraints on startups specifically in the context of solutions R&D.

Startups need a convincing story for how the project's output produce a massive and competitive return on investment over a reasonable timescale.

The long-tailed nature of startup investing means that venture capitalists need to believe that any investment in their portfolio could potentially return the fund. But it's not even enough to have great returns — you need to have the best returns! That is, VCs don't need to hit an absolute benchmark but instead need to compete against other assets like the stock market for LP (ie. *their* investors) dollars. If the stock market is absolutely crushing it, LPs have less incentive to invest in higher-risk assets like tech startups. As a result, VCs need to do everything in their power to maximize their ROI, which includes not just how much they grew the principle but how long that growth took. As a result, the same dynamic plays out 'one level lower' with VCs now in the role the LPs once held. Many atom-based technologies have the same or higher risk than a

software startup but require more capital and have longer timelines. The cold hard numbers often make SaaS businesses a better investment.

Startups need to move fast. J.C.R. Licklider kicked off work on the ARPA program that would go on to build what we would consider “modern computing” in 1962. Perhaps the first real ‘value capture’ event it led to was Apple’s IPO in 1980. This trajectory would have been untenable as a startup. While there is flexibility if a project’s valuation continues going through the roof, VC fund structure pushes for returns on 5-7 time horizons. Venture timescales often lead to projects being pushed to focus on a specific application or being acquired before achieving the technical goals they set out to hit. There is nothing wrong with focus or acquisition per se, but they can lead to technologies failing to live up to their potential.

Startups need to capture the value they create. Markets are great, but there is no law of economics that people can always capture the value they create if they are clever enough.⁷⁸ Innovation requires multiple people. Often the person who figures out how to make a lot of money off of a discovery or invention is not the person who created it, nor are they even in the same organization.

Innovations that change one or more components of a bigger system or process often make poor startups. A startup that is trying to improve or change a component of a bigger system has two options. It either needs to either convince whoever runs that system to adopt the change (which is especially hard when no particular entity runs the system) or to build the entire system themselves with the change incorporated. Obviously, the change needs to happen eventually, but many paradigm shifts (especially in complex systems) initially *lower* performance, at least on traditional metrics. Change happens either when the new paradigm has the time and resources it needs to catch up to the old way of doing things or people pay attention to new metrics where the new paradigm performs well. The latter usually only happens if they can see the new paradigm

⁷⁸ “We conclude that only a minuscule fraction of the social returns from technological advances over the 1948-2001 was captured by producers” - from Schumpeterian Profits in the American Economy - Theory and Measurement

in action and realize that the traditional metrics weren't capturing all the important features of the system. Both of these options are hard for a startup to pull off because each approach requires significant time and resources before even getting hints of traction. It's hard to get that time and resources because potential customers (the people running the systems or consuming their outputs) will assert that the idea is dumb right up until it's not.

Some innovations are just hard to productize. Process innovations often fall into this category⁷⁹ - especially those that don't depend on new technology per se, but on a new way of doing the same thing with basically the same tools. The Bessemer Process, for example, used the same crucibles but you put different materials in them. Social process changes (like [Toyota's Kaizen Method](#)) are even harder to productize. In these situations, the vast majority of the value generally accrues to the end-user. The obvious advice would be "start a firm that's doing the end-to-end thing" and outcompete the incumbents. While that should theoretically work, it often runs into the reality of extremely complex products that you would need to reinvent from the ground up (like if you had a better way of making passenger jets) or heavily regulated industries where it would take massive lobbying just to use a new process. A startup doesn't usually have the time or resources to tackle these scenarios. The alternative approach is to take on a change-management consulting role, which usually doesn't capture enough value to be VC-fundable so it needs to be profitable early on. Additionally consulting runs head-long into the bootstrapping issues we already discussed.

Startups shouldn't tackle market, channel, AND technical risk. New technological paradigms usually face all three of technical, market, and channel risk. You simultaneously need to develop the technology to a point where it's competitive with alternatives, to figure out who wants to use it, and to establish how they're going to buy it. All three are herculean tasks.

It's common wisdom among VCs that it's not a good idea to invest in companies

⁷⁹ See Fundamental Manufacturing Process Innovation Changes the World

that are taking on two, let alone three of these risks.⁸⁰ This heuristic isn't misguided. The chance of failure goes up *dramatically* when an organization tries to tackle two, let alone three of those risks. Startups should only tackle one (or *maybe* two). The reason therapeutics-focused startups can exist is that they have almost no channel or market risk; there is a well established pipeline from lab to acquisition by a pharma company and insurance companies are guaranteed to pay for drugs that go through FDA approval and target big conditions.

At the end of the day, the work you need to do to drive adoption of a technology is often very different from getting it to a certain performance level. The work priorities drive the type of organization you create. As much as we like to think that startups are about building technology, they are actually about selling technology and adopting it for specific markets so that they can do that better.

Misconceptions about things that are too engineering-heavy for academia and too researchy for startups

It's common to use wacky ideas and long-term projects as examples of things that fall outside of both academic and startup constraints. This belief is pernicious because there are many examples of both getting support; people hold those examples up as evidence that the whole class of things that don't fit into academia or startups doesn't exist.

- Long-term projects get support *if* they hit consistent milestones *and* they'll be able to capture a lot of the value they create if they succeed.
- Wacky ideas get support *as long as* they don't threaten paradigms too much *or* someone thinks they can make money off of them.
- Research projects that require a lot of people or coordination *as long as* the

⁸⁰ In addition Jerry Neumann argues convincingly in that VCs rationally should avoid technical risk in Productive Uncertainty

final product is a paper they can all put their names on.

- Research projects that require some engineering effort *as long as* the engineering effort resembles something that has been done before: building a bigger telescope / particle collider / linear accelerator / space probe, etc.

What is the relationship between profit, value capture, and impact in the context of invention and discovery?

Lurking in the firmament above any question of institutional homes for researchy activities is the question of profit and value capture. Academic work is often profit-agnostic and defaults towards being ‘open’ in a way that often clashes with capturing the value of the work. Startup work needs to be eventually-profitable so it defaults towards only that work whose value *can* be captured. Stepping outside of existing institutional structures creates a dazzling array of possible tradeoffs. Profit incentivizes people and funds work but can warp incentives and feedback loops. Value capture creates profit but can constrain diffusion and types of work. Context attenuates all of these tradeoffs!

If you want to maximize the amount of awesome, joy and wonder that technology work creates (which I’ll just refer to with the horrible but useful shorthand “impact”) what is the correct approach to profit and value capture? The topic quickly becomes ideologically saturated — some people view profit as ritually unclean and others see it as just rewards or even a key metric for virtuous actions.

I want to step orthogonally to this ideological spectrum and unpack the question “what is the relationship between profit, value capture, and impact in the context of invention and discovery?” Attacking this question isn’t just idle philosophizing; it can be a compass for navigating the idea maze by giving a sense of high-level direction while navigating around practical obstacles like

figuring out how money works at a new kind of innovation organization.

Profit is important for both practical and ideological reasons

Practically, profit increases organizational longevity and robustness because it means the organization is getting paid directly for what it does.⁸¹ In other words, profit makes organizations auto-catalyzing. Profit is the most straightforward way for an organization to both keep doing what it's doing and possibly do more of it. If the organization is doing good work, profit is the most straightforward way for them to continue!

Profit also lowers the opportunity cost of working on a project. As of 2021, most people who have the skills to do important solutions R&D work always have *much* better opportunities both in terms of pay and status. They could either start a startup and roll the dice on a massive outcome, or go work at a big tech company, get a position in finance⁸², or become a management consultant — all places where paychecks have the possibility of being huge and there really isn't any hit to prestige. Nonprofits and government organizations usually have significantly diminished pay scales because of tight budgets and convention. Anecdotally, forgone salaries is one of the biggest deterrents that keep people from becoming DARPA program managers. While many people *do* forgo higher salaries to do important work, many of them can only sustain the discomfort for so long. For other excellent people the opportunity cost is too high. Profit enables an organization to pay higher salaries, which in turn reduces opportunity cost. If more activities could be profitable, it could lower opportunity costs for excellent people to work on a more diverse set things.

⁸¹ Basically the entire world is a monetized economy as of the early 21st century. There is a whole other rabbit hole about the goodness or badness of that fact but I'm going to avoid that discussion and just treat a monetized economy as a given.

⁸² This is not a new phenomenon. Newton spent the last decades of his life working at the British Mint instead of doing physics.

Ideologically, there is a sense of justice when organizations and people who make the world better are rewarded. “Doing good is its own reward” is not untrue, but it would be lovely if profitability matched up more closely with value creation. It seems a little “off” that companies like Peloton (just to name an example and not to call them out specifically) are valued at billions of dollars in large part because of technologies like screens, GUIs, and internet connectivity while the people who originally created those technologies aren’t particularly wealthy. Anecdotally, I believe Robert Metcalfe is the only former PARC employee who became a centimillionaire and to my knowledge, no Bell Labs employees became fabulously wealthy from the work they did there.⁸³

Enabling (or at least not maligning) profit from world-improving activities is a key that unlocks both profit’s practical and ideological benefits. However, there’s a (perhaps productive) tension between whether profit should be an instrumental or intrinsic goal. Profit as an instrumental goal is the stance that “ultimately we want more amazing things and profit is an important tool to encourage people to make those things.” Profit as an intrinsic goal is the stance that “profit is the just reward for creating amazing things.” The latter can problematically bleed into the stance that “any created value can and should be captured.”

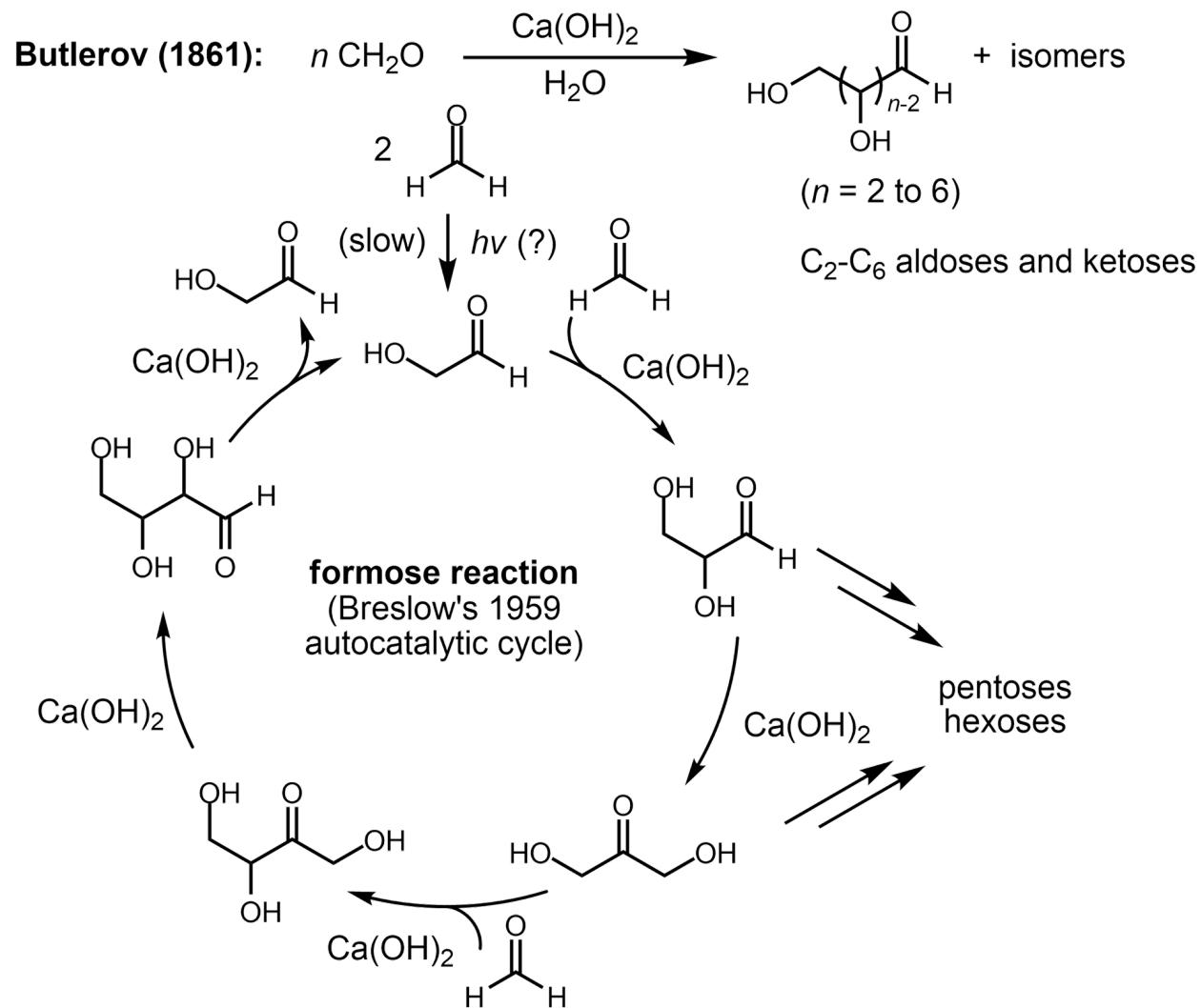
Those of us who primarily care about enabling more amazing things in the world should be wary of this last point because it comes along with the corollary “and therefore, if you *can’t* capture value, you must be missing something or it wasn’t worthwhile in the first place.” Most of you probably don’t believe the extreme corollary explicitly. However, it does tend to implicitly pervade the discourse so I’m going to spend a good chunk of time arguing against it. But before I do, I want to argue that while profit isn’t an end goal like the Golden Fleece, it’s also not a complete trap like the Hand of Midas. It’s important to acknowledge that profit is critical for both organizational longevity and for technology to diffuse

⁸³ Shannon made a killing on the stock market, but as far as anybody knows, he didn’t use information theory to do it.

into the world at all!

Profit makes organizations auto-catalyzing

Autocatalytic reactions are (chemical) reactions that produce at least one of their own reactants.



Autocatalytic reactions are complicated!

Producing their own reactant makes autocatalytic reactions easier to sustain than other reactions. Arguably life itself is one giant autocatalytic reaction. Profitable organizations are autocatalytic reactions because they produce the key ingredient you need to keep an organization going: money. In effect, profitable

organizations become their own money factory. Autocatalytic organizations are more robust and ultimately have more leeway to take unconventional actions.

Profit gives an organization as much control of its own fate as possible. A profitable business has many different ‘moves’ at its disposal that it can use to survive. It can build up cash reserves to hold it through down-markets; it can sell stock to raise cash; it can shift products to adjust to changing demand; it can grow or shrink depending on the need. All of these moves are possible with minimal regulatory restriction or third party permission. These moves can increase organizational longevity which in turn can enable an organization to take on long-term projects that would otherwise be off the table. In addition, larger profit margins translate to more slack in the system for unconventional actions. Low margin organizations (rationally) avoid rocking the boat because they can quickly become unprofitable while high margin organizations can afford failed experiments.

On the other hand, if you are funding operations from sources beside profit you’re obligated to one or more entities that care about something besides the organization’s output. Some of these sources include purchasing debt, selling equity, or soliciting donations. Ideally, incentives are aligned even in this situation, but ‘reaper functions’ — the real reasons, often unspoken, why people are promoted or fired —dominate other incentives. At the end of the day, equity investors care about their share prices going up, debtors care about being paid back, and donors care about feeling good.

The only way for an organization to become auto-catalyzing without being profitable would be to somehow amass a war-chest large enough to fund operations out of interest on its principal. Looking at a list of the longest-lived organizations in the world, they’re overwhelmingly profitable companies that make and sell goods people want (usually food or alcohol) and universities with endowments.⁸⁴

On top of flexibility and long-term survivability, profit also creates an extremely

⁸⁴ See The Data of Long-lived Institutions

tight feedback loop between an organization and the outside world.⁸⁵ Arguably, cold, hard cash is one of the only metrics that isn't automatically subject to Goodhart's Law⁸⁶ — if your organization's goal is genuinely selling a product or service, profit is a hard metric to game short of literal fraud.

While it's nowhere near as useful as actual profit, aiming for eventual profitability can still contribute to organizational longevity. Practically, it's both easier and less organizationally risky to raise capital if you're aiming to become (or better yet, are) profitable. Many people are more open to parting with their money if they think of it as an investment instead of a donation. And obviously the less uncertain and closer to the present that profit is, the less you need to pay for that money. In the context of talking to people about riffing on ARPA I've gotten "can I invest?" many times, but rarely "can I donate?" Of course, targeting eventual profitability can also backfire from an organizational longevity point of view!

The trick is threading the needle between organizational goals and what people will actually pay for. We tend to laud organizations that have pulled this off — they become household names like Apple, SpaceX, Zappos, etc. It's pretty clear that they are better equipped to deliver delightful electronics, spaceflight, and internet shoes than a nonprofit with the same goals. Alignment is easy if your organizational goal is just 'make as much money as possible' but that is often not the case for organizations in the 21st century that are often started to fulfill some mission "and to make money while doing it." Most of you reading this probably lean towards the latter category. I suspect that an underlying point of disagreement between people who laud profit and those who denounce it is over how feasible it is to align organizational goals and profitability. It's easy to get sucked to one end of the spectrum or the other: either the view that it's possible to align profit and other goals in almost every situation or that profit inevitably distracts from other goals. Like most dichotomies in complex systems, the

⁸⁵ Jason Crawford argues that cash-flow also creates the tightest possible feedback loop with the world outside the organization.

⁸⁶ "When a measure becomes a target, it ceases to be a good measure."

answer is “it depends.”

Someone needs to buy manufactured technology eventually in a monetized economy

Profit is also an integral part of getting technology out into the world. The only way that a technology ends up being used by people who didn’t build it is if it’s sold or given away. The former situation is far more common. In either situation, technology needs to be ‘productized’ ie. made legible and useful for people outside the organization. If it is going to be sold on an open market, it also needs to be ‘commercialized,’ ie. made cheap and desirable enough that it makes sense to sell it in the first place. Empirically, profit-seeking organizations have much stronger incentives to do commercialization well. As a result, profit is almost always part of technology out into the world.

There are exceptions, of course. The major exception happens when a technology is created and used entirely internally to an organization. Many of the internal tools used at Google are entirely created, manufactured, and used internally. Other exceptions, like open-source software, are harder to categorize. Firms create technology barter deals with each other. At-home manufacturing completely breaks the model, although arguably that would just be a trend towards a less monetized economy.

It’s a common perception that the military creates internal technology similar to Google and sits outside the market as an unmoved mover of technology. The US government’s version of internal technology creation is less clean than Google’s. The government does own R&D labs it uses to create some technology that it specifically needs. However the government contracts out much of its R&D and all of its manufacturing to external firms (as opposed to owning the manufacturing plant and giving it a budget.)⁸⁷ So even military technology is

⁸⁷ To complicate things, many small contracting firms only work for the US government so they may be part of the government for all intents and purposes.

purchased eventually. The firms that sell to the military need to pay *their* bills so the costs do matter (though cost-plus contracting distorts this) and if the military paid *less* than the cost of the goods the firms would go out of business. And while the US military spends hundreds of billions of dollars a year (\$934B in 2019), it does not have infinite money. So while they are weaker, I'd argue that the forces of market discipline do act on military technology - if it becomes cheaper for the same quality it will be used more.

If this view is correct, it means that in a monetized economy, commercialization is an *essential* step on a technology's path to adoption. There is an implicit feeling among some technologists⁸⁸ that commercialization is a lesser activity for people just out to make money. Instead, it should be seen as a critical part of the cycle.

Market discipline is a double edged sword

And of course, markets are where you buy and sell goods. Markets are implicitly entangled in any line of thought around commercialization and profit.

The most interesting thing about markets for thinking about enabling better solutions R&D (and invention+discovery more generally) is *market discipline*. I'll be using it in the colloquial nebulous sense of "she wanted to build an entire new computing ecosystem but market discipline forced her to release subpar imitation of existing products"; "resisting market discipline, she built the technology everybody thought was stupid but turned out amazing in the end"; "market discipline honed the organization into a lean mean product producing machine" rather than the narrow, technical definition that nobody actually knows. Market discipline can make organizations strong and efficient, but it can also crush eventually-productive creativity. Discipline enabled the Spartans to crush the Athenians in the Peloponnesian War, but do we want a world made of only Spartas?

⁸⁸ My past self included.

Decoupling from market discipline is like cave diving

Note that I have never gone cave diving. I am not that cool.

Market discipline is embodied in Y Combinator's motto: "make something people want." It's a good heuristic for hewing to market discipline even before making money. Paradoxically, there are many counter-aphorisms like the apocryphal Henry Ford quote "if I asked people what they wanted they would have said a 'a faster horse.'" The resolution to this paradox is "Make something people want at some point in the future." In the long run, someone needs to buy something downstream of any 'useful' work, so by definition they need to want it! However, the whole time you're *not* making something people want, you are burning resources. The question then becomes "what is the thing people want in the future and can the technology or idea survive long enough to get there?"

Cave diving feels like a great analogy for the process. While you're going through the cave, you're using oxygen and if you run out before you reach the next air pocket, you're dead. Working on a product you can always test against the market is like normal diving - you can always come up from air and maybe you need to get the timing right so you don't get the bends.

Perhaps it's torturing the analogy but there are some interesting concepts to pull back from cave diving. The longer the path, the better you need to plan. If it's a short hop and you can see where you're aiming you don't need to plan as much. If you know that there are lots of possible air pockets, you have the freedom to explore more. Similarly if the technology has lots of market 'exit ramps' you can do more exploration. Software is uniquely repurposable which might offer an explanation for why agile product development makes sense for software but perhaps not in other situations. Having a precise idea of where you're going is very important. Even if you don't know the exact way, you need to keep your bearings towards where you're going.

Many of DARPA's outlier results were things that nobody was asking for

DARPA works on programs that go against the established paradigm and often people who have become comfortable with one way of doing something are not asking for you to come in and change it. DARPA has worked on things un-asked-for by both the military and industry.

Three examples of this are:

1. **Drones** — DARPA had a drone program, transferred it to the navy where it was discontinued, but DARPA continued to work on them until the military paradigm shifted and suddenly they were incredibly useful.
2. **Optoelectronics** — in this case it was businesses that weren't asking for optical multiplexing. AT&T and IBM had worked on multiplexing but had budget crunches and didn't see value in it. DARPA supported optoelectronics work from 1985->2005 (20 years!)
3. **Personal Computers** — nobody was asking for most of the pieces of the Mother of All Demos.⁸⁹ Instead of collaborative shopping lists and voice calls, the military wanted a computer that could help commanders understand what was happening with nukes and battles.

The upshot is that you can't apply the normal maxim of "Make something people want" to an ARPA program and seek out "customer validation." This missing feedback loop is tricky because you do need to make something people want ... eventually.

⁸⁹ The Mother of All Demos was Douglas Engelbart's tour-de-force where he demonstrated the majority of what has become modern computing in 1968.

Profit-maximization can hamstring technologies whose impact depends on them

Profit-focused organizations are incentivized to try to capture as much of the value they create as possible. At the end of the day both a startup's valuation and a corporation's share price are both theoretically based on future cash flows and profits. Of course, profit-maximizing organizations are not mindless greed-monsters. In many situations, giving up short-term profit can actually lead to longer term revenue. Ideally, these organizations would be able to sit at some point on the tradeoff curve that maximized value *by* maximizing impact. Profit-maximizing organizational structures can also be the best way to support a technology!

However, there are four mechanisms that can lead to for-profit companies hamstringing technological advancement, despite other aspirations.

1. Specialization — Generally, it's a good idea for a for-profit business to start with a single product and a specific market or niche. This good business practice means that outside of a large company with an R&D department, most of the development work should be focused on productizing a technology for a specific task. Startups who want to start by building a platform or *doing all the things* tend to go down in glory-seeking flames. Specializing technology for a specific task is not inherently a bad thing! It's also possible to thread the narrow path of increasing generality, starting from a small niche and jumping to bigger and bigger applications over time. Elon Musk is the master of this. However, companies rarely avoid getting stuck in one of those niches, especially if the niche isn't profitable enough to support R&D on the more general technology. This situation happens all the time to university spin-off technology; it often needs more general development to reach a certain performance threshold, but it gets stuck in a niche in order to build a company. I've seen this trap ensnare everything from multi-material 3D printers to biosensors to home robots.

2. Investor Pressures — For-profit companies working on technology generally need investors because they won't turn a profit for at least several years after

getting started. These investors have their own incentives, and maximizing the impact of a technology generally is not the most important one. Investors can have specific timelines for an exit, need to show growth to *their* investors in order to raise a new round, or just opinions about where the technology should go that conflict with maximizing impact. (Of course, any source of money comes along with its own set of pressures!)

3. Capturing value — For-profit companies need to capture some value that they create! It is rare that an organization can both forgo significant value by making key parts of a technology public *and* make money at a rate that both satisfies investors and can fund the development of general-purpose technologies. I will argue later that there is a significant class of innovations that would create drastically less value for the world if their value had been captured by their creators.

4. Speed — An organization could detect all of the traps listed above and forgo trying to maximize profit, not take investment, and make most of its technology work public, pulling in revenue through non-exclusive license fees and consulting work. Unfortunately, this strategy can hamstring the technology you're working on in a different way, limiting general-purpose development to the work that can be done when you have particularly lucrative contracts. As a result, technology development slows to a crawl. Many aerospace technologies seem to stuck in this particular trap!

We stand on the shoulders of too many giants to give them all credit

If you wish to make an apple pie from scratch, you must first invent the universe.

— Carl Sagan

Should the neighbor who had a cross-fence conversation about a process she uses at work that sparked the realization that led to the invention capture some value?

She was certainly instrumental to its success. What about the technician who knew how to twist the wire in just the right way to get the first proof-of-concept to work?

And what about all the failed projects that left a pile of skeletons indicating where the traps were?

It might be possible for software companies to go through their tech stacks and enumerate the open source projects that they use and who contributed to those, there is still a ton of illegibility around what fraction of the value each project contributes, how much each contributor to the project contributed, etc. And outside of software it gets even more absurdly gnarly.

The multiplicity of people involved in creation leads to a contradiction that I suspect is fundamentally unresolvable:

On the one hand one person or class of people do not deserve credit for an innovation. The iPhone needed Jobs *and* Ive *and* all the engineers and designers at Apple *and* the team at Xerox PARC *and* all the people who invented the transistor, and all the folks at Intel and other places who made it cheap *and* everybody at Corning who made gorilla glass *and*

On the other hand credit is really important for motivating people and enabling us to wrap our human brains around the world.

Current value capture mechanisms are crude

The direct ways to capture the value from research are to patent⁹⁰ it or build a product around it (and leverage trade secrets.) Other possibilities are more indirect and ‘credit based’: consulting, speaking fees, or leveraging reputation to get more lucrative positions later.

⁹⁰ I'm going to just talk about patents but the same arguments apply to both

Patents

Patents require an expensive application that can take years to process, lawsuits to enforce, and effort to monitor for violation, and lead to an extremely illiquid market. On top of that, there is tons of valuable work that just isn't patentable.

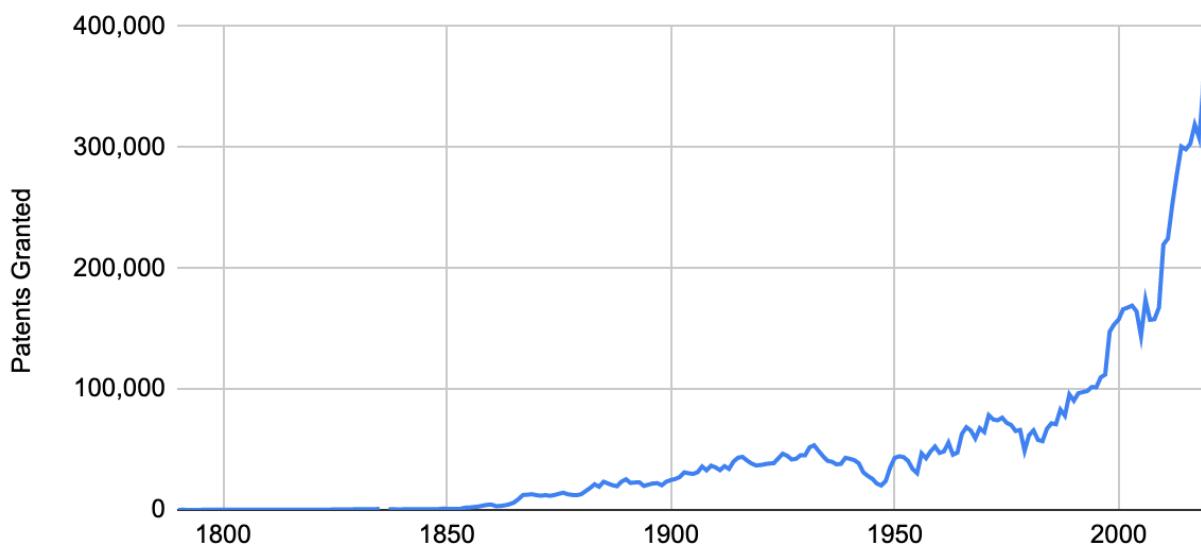
Capturing value with a patent is often inefficient. In an ideal situation, someone building a product that incorporates the patented process comes to you, the two of you negotiate a licensing deal, and they send you a check every year. The world is not ideal. Often, you need to *find* people who are violating your patent and threaten to sue them if they don't pay up. Standard lawyer advice when building something is to *not* look for prior art because it decreases your liability in case you happen to reinvent something. Even if someone *does* come to license your patent, the negotiation over licensing terms can be brutal - there are no standard terms and especially for a startup it can create a drag chute on a product's profitability that makes an already hard job even harder. Once the licensing agreement is signed, you still need to make sure that the company is paying you the royalties you're due and sue if they're skimping you. Many inventors have spent a lot of their lives in court over patents and sometimes wind up dying in the gutter in spite of that.

It's worth noting two historical facts about patents that shed some light on their limitations. Patents have gone from well-reviewed descriptions of already-built mechanical inventions to hastily-reviewed descriptions of everything from software to biotech written by lawyers. The number of patents issued in the US has increased by four orders of magnitude since the patent office opened — from 33 in 1791 to 354,430 in 2019.⁹¹ The number of patent reviewers hasn't increased anywhere near that much. On top of all of that, patents weren't even created as a value-capture tool for discovery or invention. Instead, patents started as a way for monarchs to increase their income without going through Parliament by granting temporary monopolies.⁹² Of course, the modern patent system is

⁹¹ This is two orders of magnitude larger than US population growth.

⁹² See Age of Invention: How to Build a State

explicitly meant to incentivize innovation so this might be nothing more than a historical anecdote.



Additionally, patents inherently prevent a piece of intellectual property from being combined with other knowledge. The combination-preventing nature of patents is particularly problematic if you want to use them to try to fund work to address gaps in work on processes and systems (which naturally require technological combinations).

Products

The first reason products are a crude way to capture value from research is that most modern products are an amalgamation of different technologies and designs. So while a piece of research may have vastly improved a component, that doesn't necessarily translate to a vastly superior product.

Second, products are inherently coupled to companies, so the success of a product (from a value capture standpoint) is inherently tied to the success of a company. This coupling means that there are many other factors that affect the product's success besides the product itself (let alone the research-based piece of it). Markets, sales channels, margins, fundraising, perception, etc. all make or break companies independent of their products. "The best technologies don't always win" is an adage for a reason.

“But unique new technology creates a moat that will make the company successful!” Jerry Neumann’s [Productive Uncertainty](#) argues convincingly against pure-technology with the idea of “excess value” an invention would produce for a startup.

If a moat exists prior to the startup being founded (say, a patent) then, absent uncertainty, this patent could be sold for at least as much as the startup could garner from it.

More broadly, product-driven value capture forces innovations to be subject to the constraints on startups or corporations. None of this is to say that products are a *bad* way to capture value, only that they are subject to many constraints and contingent requirements that mean that the correspondence between valuable technologies and technologies whose value can be captured by a product is weak.

Patents inherently prevent a piece of intellectual property from being combined with other knowledge

Imagine a system with components A, B, and C, each of them owned by a different entity. On their own, they are worth one unit but the system is worth ten units. Arguably, the value of the last unit added to the system is actually eight. A game theoretical situation occurs where each IP owner holds out to be paid eight units. An alternative situation is that the value of the system is five but A, B, and C all arbitrarily set their price to be two.

This situation is exacerbated the more different applications there are.

This combinatorial drag from patents is why In most industries outside of

pharma and chemicals, patents are a net drag on innovation.⁹³ Chemicals and drugs usually have a 1:1 correspondence between patents and products. On the other hand, semiconductors and software can have dozens of patents in a single product.

What about a non-exclusive license? If it's under a non-exclusive license that effectively becomes a public good. Non-exclusive licenses have other problems - they enable you to capture less value and still require lawsuits and monitoring to enforce.

Value capture can hamstring impact

We often talk about individuals or organizations capturing the value that they create. However, conversations about value capture can shift from an instrumental goal in service of encouraging more innovation to an intrinsic goal: that people should be justly rewarded for the value that they create. I want to argue that from a practical perspective, value capture as an intrinsic goal is a chimera and point out its significant downsides.

Value capture as an intrinsic goal requires that you (or at least a market) can put a price on everything, and that just isn't true. Norman Borlaug arguably saved a billion lives. What is the value of those lives? The people who benefitted most from his work were some of the poorest people in the world, so their contribution to GDP is negligible. And yet, their lives have value. What is the value of knowledge about how the universe started, or the weird blob-fish that live at the bottom of the ocean? One could make an argument that perhaps in the far future our studies of the Big Bang will lead to warp drives, or the blob fish will lead to a cure for aging. I can't shake the belief that this knowledge is important even without contributing one cent to GDP and any mechanism to find a "dollar equivalent" would be absurd. Markets are amazing, but they

⁹³ For a much deeper discussion about where patents are helpful or harmful see The Patent Crisis and How the Courts Can Solve It

require certain conditions to work and it's not clear that those conditions can be met for everything.



Belief in value capture as an intrinsic goal warps people's focus. If you believe that value capture is an intrinsic goal, it makes sense to focus on value capture mechanisms without consideration to whether those mechanisms are the most effective way to encourage valuable behavior. Viewing value capture as intrinsically good encourages the stance that if someone or something is *unable* to capture value, it must not have been valuable in the first place. If you asked, many people don't explicitly believe this, but pay attention and you will see it implicitly worm its way into conversations and logic.

Characterizing creations hamstrung by value capture

Two notes

1. By 'public' I mean technology that is not illegal to copy without permission. It doesn't necessarily mean it's 'open' (though that does make it more public.) It also doesn't mean that it was funded by the government (which incidentally funds plenty of private inventions.)
2. To avoid introducing a new term I'll be using the term 'creation' to

describe nebulous discovery/invention/potential innovation type things.⁹⁴

How can we think about the nebulous line between creations that would hamstrung by value-capture and those accelerated by privatization?

You can approach the question both empirically and theoretically. Both approaches are flawed. Empirically there are anecdotes to support any position. Theoretically you can make arguments on both sides and the ultimate conclusion is something around “it depends.” I’m going to focus on the theoretical approach to outline the characteristics of creations that would be hamstrung by value capture and lean on the empirical approach to argue that potential innovations that have these characteristics are increasingly common in the world.⁹⁵

Ignoring incentives and costs, scientific and technological knowledge will have more impact if it is more public. If knowledge is more public and accessible, more people can try more experiments, and more brains will be able to focus on the real problems. The creation of technology is a combinatorial process⁹⁶ and public knowledge enables more combinations.

That’s a big ‘if.’ Incentives do exist and research can be expensive.

So there are good arguments both for holding knowledge private and for making it public. We neither live in a socialist paradise ruled by an omniscient overlord who distributes resources to their optimal use nor a world where frictionless markets exist for everything, distributing value perfectly to its creators. So the answer to “will a creation have more impact if its private or public?” is “it depends.” Our job is to figure out what makes the answer tip to one side vs. another.

⁹⁴ ‘Heuristic’ is a better word, but I’ve introduced enough new terms as it is.

⁹⁵ This section only skims the surface of this topic as far as we need to understand the sorts of technologies that current institutions have trouble enabling. For a much deeper dive see the already-mentioned-but-worth-mentioning-again The Patent Crisis and How the Courts Can Solve It

⁹⁶ See The Nature of Technology: What it is and How It Evolves and One Process

One important question is whether a technology is valuable on its own or in combination with other technologies. And, in the latter case, how many other technologies? Single-molecule drugs are a great example of low-combination technologies and airplanes are great examples of high-combination technologies. Patents inherently prevent a piece of intellectual property from being combined with other knowledge so patents will be problematic if the majority of a creation's value will be generated combinatorially. You could think of patents as a tax on including a component technology; taxing a thing is generally a good way to get less of it, all things held equal. If there are only a few big, clearly valuable possible combinations, IP (especially a non-exclusive license) may still be reasonable. Patent's combination-taxing mechanism suggests that 'many combinatorial possibilities' may be one heuristic for creations that will have much more impact if they're public.

There is also the question of composability⁹⁷ — how easy is it to do the combining? How large is a technology's 'combinatorial surface area' and how much do you need to mess with its insides to combine it with other things? If a creation has a small surface area and / or is very composable it can have a lot of impact as a private product. Take bolts for example. They are *only* valuable when you combine them with other things to create something new. However, they have a small combinatorial surface area (you almost always use them to hold things against a threaded thing) and you rarely need to modify them (if you do, you're probably doing something wrong but I have an angle grinder to sell you). Well-designed APIs are also extremely composable. Low surface-area, composable creations make good private products that can achieve near-maximum impact.

Inverting the characteristics of creations that *can* create significant impact as private goods is suggestive of the sorts of things that would create significantly less impact if you tried to capture their value: creations that generate uncertain value through combination with many other things, have a large combinatorial

⁹⁷ Composability is an excellent and under discussed concept. <https://en.wikipedia.org/wiki/Composability>

surface area, and frequently require internal modifications to enable these combinations to happen.

How common are potentially hamstrung creations?

Even if there's a class of creations whose impact is stifled by trying to capture their value, the question remains: "how common are these creations, actually?" People trying to capture the value from their creations have generated *massive* value in the world, so maybe anything else is a corner case. I will argue that they have been common throughout the 20th century and that increasingly complex technology makes them even more common.

Caveat: There are two glaring problems with this incredibly empirical approach. First, we cannot go through *every* technology, so there will be inevitable cherry-picking. Second, counterfactuals are hard. There is no way to 'prove' that a technology would have had more or less impact if it were public or private. With these problems in mind, the empirical approach is probably most useful to try to *disprove* the assertion that there is a non-trivial set of heuristics whose impact depends on being public. If the disproof feels weak, that is perhaps the best empirical evidence for the claim.

The innovation ecosystem of the late 19th and early 20th century is one of the strongest pieces of evidence that private innovations alone can carry the torch of progress forward. It was an age of massive technological change — vulcanized rubber! Automobiles! Airplanes! Electricity! Telephones! Almost all created by private for-profit companies, patented, and sold.

However, many of these innovations only hit the mainstream when the component technologies became public knowledge. There was an explosion of car patents in the 1880s and 1890s and an explosion of car *production* in the 1900s and 1910s, roughly tracking the end of a patent's 20 year lifespan. It's telling that safety glass was patented in 1905 and became standard car equipment in 1926. Of

course, this could be completely correlational and the time gap between invention and diffusion is just how much time was needed for learning-by-doing to happen.

None of these technologies were complex enough that you couldn't figure out how it works by taking it apart. Nor did they require up-front capital costs inaccessible to individuals without institutional backing. Those attributes mean that almost by default the technologies became public knowledge as soon as they were sold. The sufficiently low cost and complexity combined with the cost of acquiring and enforcing patents in multiple countries at once⁹⁸ meant that while private, for-profit organizations were creating the technology, the technology itself was effectively public. It was virtually impossible for a single organization to blunt its impact across the whole world. The t

Many of the 'charismatic'⁹⁹ technologies of the late 19th century were 'final goods' with relatively small functional surface area. Cars move you and your stuff fast, telephones let you talk to someone else, airplanes move stuff through the air. Even electricity was primarily just used to light stuff up and spin things. Additionally, you could make these technologies good enough to have *some* market using relatively low-complexity component parts that could be created in-house. In situations where those component parts were private thanks to patents or just secrets, the invention's impact was suppressed until the components became more public. The canonical example is the Wright brothers' patent on coupled wing-warping / rudder control for airplanes¹⁰⁰ which arguably¹⁰¹ stunted the American airplane industry for a decade.

The upshot is that few Industrial Age technologies meet the criteria that we

⁹⁸ It was (and still is) a giant cost and pain.

⁹⁹ In the sense of 'charismatic megafauna'

¹⁰⁰ See https://en.wikipedia.org/wiki/Wright_brothers_patent_war

¹⁰¹ There literally is an academic argument over this. For arguments that patent thickets did stifle airplane development see Spillover Effects of Intellectual Property Protection in the Interwar Aircraft Industry and for arguments that patents weren't a problem see The myth of the early aviation patent hold-up—how a US government monopsony commandeered pioneer airplane patents. I find the former compelling but it's tricky.

would predict should lead to a technology being hamstrung by being private. Strictly enforced patents on the technologies that approach the criteria (like airplane control) seem to have hamstrung impact in the US where the patent was in effect.

So which technologies *would* meet the criteria for being hamstrung by attempted value capture? The age of chemical engineering that began in the early 20th century seems like good hunting ground. Chemicals and the scalable processes to make them are rarely a final product and can be tuned for many different purposes.

The Haber—Bosch Process is one of the most charismatic chemical processes. It also turns out to have been made public through a non-exclusive license at gunpoint. From [Enriching the Earth](#):

The first transfer of the Haber–Bosch process abroad was a result of the Versailles Treaty, which the defeated Germany had to sign in 1919. By its terms, BASF was obliged to license construction of an ammonia plant with an annual capacity of 100,000 t in France.

It would be an entire (worthy!) research project to discern the licensing practices around other chemical processes. However, skimming through the history of the heavy hitters there seem to be repeated references to multiple companies bringing them to market.¹⁰² Additionally, a significant chunk of the work still ended up in academic papers instead of patents, whether it was being done in universities or industrial labs.¹⁰³ The number of papers that industrial labs published suggests that at least more of the steps leading up to an innovation were public than today.

It's worth noting that before the relationship of government, academia, and technology fundamentally changed in 1980 with the Bayh—Dole act, it was hard

¹⁰² See Not Counting Chemistry: How We Misread the History of 20th-Century Science and Technology

¹⁰³ See The Decline of Science in Corporate R&D (again)

to have exclusive license rights on federally funded research. The situation flipped after 1980 to a world where research institutions by default had exclusive rights to the output of federally funded research. This shift coincides roughly with the distinct absence of new general-purpose physical technologies. Obviously any direct causal link is speculation at best. At the same time, ignoring it feels dishonest as well. Counterfactuals are hard.

The computer age is littered with anecdotes of technology that was only impactful because its creators failed to capture its value. Alan Kay argues that the technologies created in Xerox PARC have created trillions of dollars in value.¹⁰⁴ Imagine a world where Xerox sued the living daylights out of any product that used a mouse or a GUI. I suspect the mouse and GUI would have been much less impactful in that world.

Counterfactually, what would have happened if AT&T were not required to license for free all of the non-telephone patents at Bell Labs? Two plausible scenarios could have happened: Either AT&T would have had a monopoly on the semiconductor market or William Shockley would have had an exclusive license to the technology. In the former situation, it's likely that semiconductors would have been treated the same way as most technologies tangential to a core line of business historically have been: marginalized and deprived of the chance to live up to its promise. An eternally expensive piece of military equipment. If instead Shockley negotiated an exclusive license with AT&T, consider that Shockley was a notoriously terrible manager and Shockley Semiconductor went out of business in 1968. The traitorous eight would never have left to found Fairchild, so Shockley Semiconductor might have persisted, but would the innovations that came out of Fairchild have happened? Would Hoerni's Planar Process have been created and adopted in either situation? Poor management is great at killing weird paradigm-shifting ideas. The entire modern technological stack might never have existed.

Linux is literally powering Mars Rovers now. How much irreplaceable impact

¹⁰⁴ See This Interview with Alan Kay

has it had? If Linus Torvald were trying to capture the value it created, how would Linux have been anything other than an inferior version of windows?

Conclusion

Many ideas have massive impact *because* people can copy them without paying their creator. Any mechanism for capturing value has more or less built-in friction. This friction can limit the diffusion of innovations.

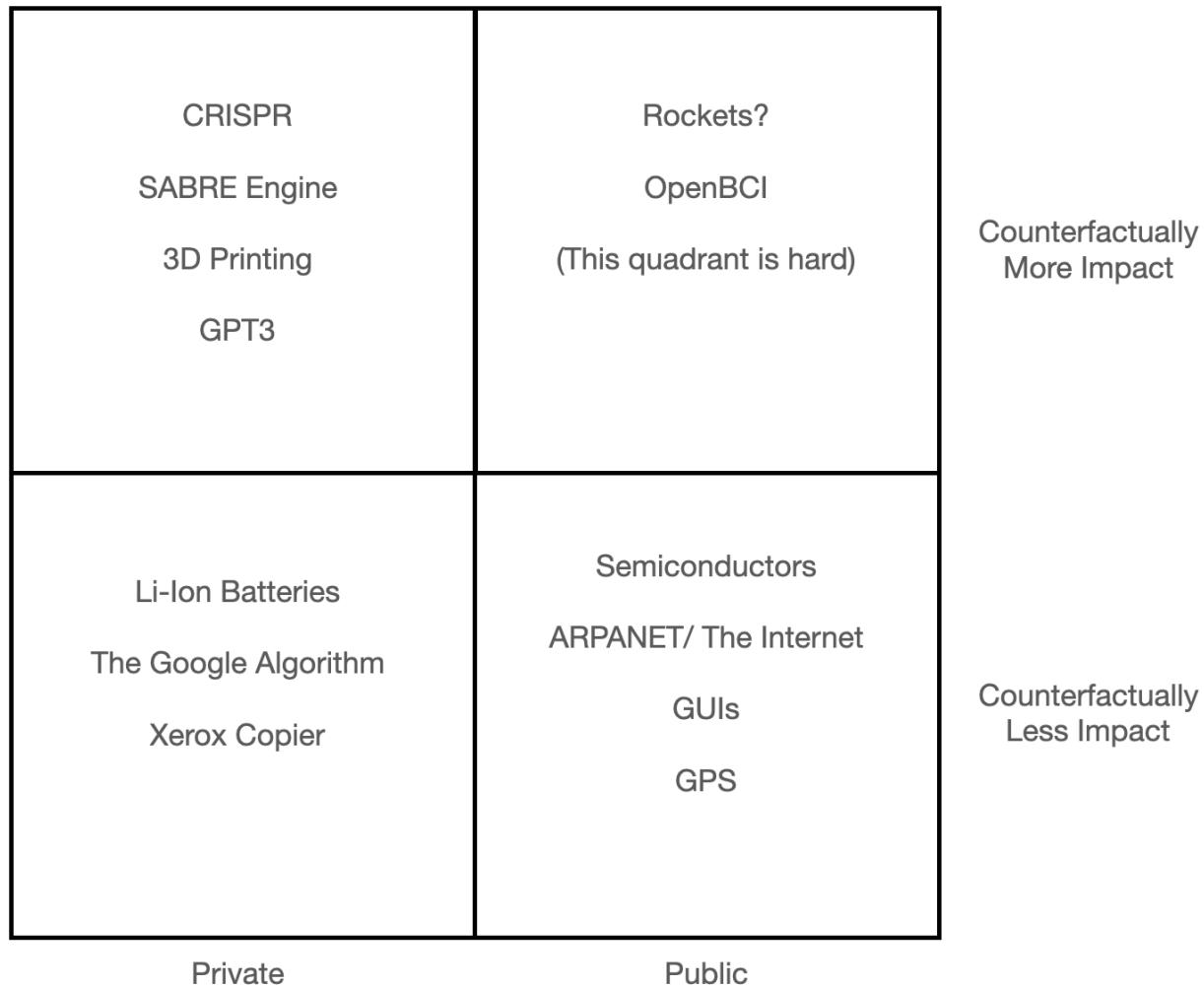
The strong argument in favor of trying to capture value is that value capture is what incentivizes and funds innovations in the first place. Value capture is not a binary “either you capture all of the value or none of it” but a question of how much of it you capture and how you capture it. Arguably Google (and many other companies) have created much more value than they have captured and may have created *less* value if its creators hadn’t tried to capture that value. Profit really is important! However, these companies generally are not generating profit from innovations with large surface area that generate uncertain a priori value through not-particularly modular mechanisms.

Current value capture mechanisms are crude. It’s possible that better value capture mechanisms could create a world where any innovation could create as much value as possible *and* its creators could be rewarded. However, it doesn’t seem possible to capture value (especially in non-digital technologies) without introducing *some* friction into the system in order to track contributions and forcibly extract money from people who are keeping more than fair share¹⁰⁵.

Finally, the potentially-hamstrung class of innovations seems to become more prevalent as the complexity of technology increases and technological development depends more and more on long research pipelines. Given all of these considerations it seems clear that navigating between impact and value capture is a Scylla and Charybdis situation indeed. Like other unavoidable tensions, there’s no ‘correct’ answer beyond “constantly pay attention and course correct!”

¹⁰⁵ The equivalent of suing someone who isn’t paying the licensing fees on a patent.

For fun, I will leave you with a potentially controversial 2x2 of innovations that would arguably have had more or less impact if they were public or private.



Legal structure is critical to institutional capacity

While we didn't mention them directly, legal structures were lurking just below the surface throughout the discussion of profit and value capture. *Who* exactly is capturing value and *how* profit incentivizes individuals is all a matter of legal structure. Many people implicitly or explicitly believe that legal structures are akin to "syntactic sugar" in programming: a surface-level feature that ultimately has little effect on organizational outcomes. I want to argue that instead, legal

structures are intimately tied to organizational capacity.

When you boil it down, legal structures are contracts that establish how money and explicit power flow in an organization. Served that way, how can legal structures not be tightly coupled to organizational capacity? The government acts as an external forcing function on these contracts: allowing some, disallowing others, and using them to assess one of the two inevitabilities of life (taxes) which actually play a big role in what people decide to do with their money. Taxes often play a big role in incentives for people and organizations with enough money to make a difference with their money.

Historically, new legal structures have enabled new classes of activity. The creation of joint stock companies in the 1600's enabled larger and riskier private ventures like the East India Company¹⁰⁶ and arguably enabled a lot of English settlement in North America. The creation of limited liability corporations in the mid 19th century helped enable the railroad boom first in Britain and then in the US.¹⁰⁷ In 2020, SPACs made going public a reasonable move for several non-software technology companies that may not have happened otherwise.

[Entrepreneur First](#) is a great example of an organization using unconventional structures to unlock new strategic moves and outcomes. They are in part an investment firm, but their cohorts require much more operating capital (ie. more employees) than the traditional VC firm structure provides. So, in addition to the normal-for-VCs closed-end funds with LPs, EF also has a C-corp that issues equity and has its own investors that both employs most of the people at EF and "co-invests" with the closed end fund. This unconventional structure allows EF to be more operationally intensive than a normal VC fund and has the added side effect that if their investments are successful enough, it's conceivable that they could choose to stop needing LPs at all. Eventually the organization could become a self-catalyzing cycle where operations are funded solely through liquidated investments held by the C-Corp. [Y Combinator](#) has a similar structure

¹⁰⁶ See A Brief History of the Corporate Form and Why it Matters

¹⁰⁷ See The Modern Corporation and Private Property.

and uses this additional flexibility to do experiments like the now-defunct [YC Research](#). This additional long-term operating flexibility wouldn't be possible in a normal VC fund structure because most VC firms fund operations off of their management fees and return all their profits to LPs. I want to make it clear that there are tradeoffs so this is not a strictly-better structure for all tech investors. But it is a better structure *for EF and YC* to the extent that they conceivably could not do what they do without it.

So if new legal structures have a large impact on organizational capabilities, why shouldn't organizations trying to do qualitatively different things take legal structures seriously? Clearly there are many organizations that have accomplished great new things without innovating on legal structures. This is yet another place where counterfactuals are hard. We do more than speculate on organizations that structural constraints caused to never exist in the first place. Other organizations may have been subtly forced into more main-stream behavior. However, the historical precedent and surprising number of ways you can be creative with legal structures suggests that, at least on the margin, there is a lot of room for new legal structures to enable new organizational capacities.

Legal structures are intimately tied to the question of how money works in an organization

Legal structures affect where you can get money and what you can do with it. Even the words we use have both a funding and an organizational meaning. “Nonprofit” both implies that funding is from people who don’t expect a return on profit and is a specific designation in the law. Similarly “selling equity” implies a specific legal structure and also that funding will come with a certain set of expectations.

“How money works in an organization” spills over into organizational incentives in too many subtle ways to list. For example, creating equity enables you to raise money that doesn’t need to be returned on a finite timescale (which can

incentivize longer term thinking) but puts ownership of a fraction of the organization in the hands of outside shareholders (who can push for different decisions than the founders would have made.¹⁰⁸) If you want to avoid someone else owning your organization but still want to raise money from outside investors, you could organize as an LLC but then you start a ticking clock to return that money with interest. Nonprofit status enables you to raise money that people can write off on their taxes and don't expect a return on, but comes with a host of requirements around what you can and must spend money on. These examples are only meant to illustratively scratch the surface — someone should really write *The Big Book of Organizational Incentives*.

As an example of the coupling, let's think through funding a DARPA-riff. It is hard to capture value from research so a reasonable first instinct is to look to philanthropic funding. Philanthropic funding pushes an organization to legally be a nonprofit both because people expect a tax break for money they give away and 'nonprofit' signals "no seriously we're not going to use your money to get rich."¹⁰⁹ That's not the end of the story though, because from the perspective of pure impact the best move will likely for a startup to carry some of the technology forward. It would be extremely helpful to use some of the startup's eventual-profit both to fund further work at the DARPA-riff and to tap into for-profit investment. People are willing to write much bigger checks if they think they'll get a return on them! However, the standard nonprofit structure puts many constraints on your ability to issue equity (which would be the normal structural way to channel for-profit investment) ...

¹⁰⁸ Note that this isn't a good or a bad thing - while the default reaction at least in Silicon Valley circles is that it is.

¹⁰⁹ Although cognitive-dissonance-introducingly there are nonprofits that make a large profit in everything but the legal sense, and for-profits that don't seek to maximize profits at all.

The traditional nonprofit structure is a bad idea for an organization that requires a significant chunk of capital and produces uncertain results on a long timescale

A non profit can be organized as either a charity or a foundation. Each has different constraints, but in both cases, Nonprofit structures are more at the whim of the entities giving them money than if they were providing a good or service. It just comes down to leverage.

Legally designated charities enable donors to write off their donations as tax deductible. However, charities need to get 33% of their income from donations and those donations need to come from 'small donors' who are each provide less than 2% of their total funding. Because of these rules, charities raise money on a yearly or sub-yearly basis which forces them to give people a reason to give on that short timescale. Uncertain results on a long timescale don't lend themselves to the sorts of updates that encourage people to give on short timescales. "Hey, uh, still doing that experiment that we were doing last time we asked you for money." The short funding timescale also means that cash flows can be quite volatile, which poses obvious problems for long-term research. You can't really fire a researcher and then hire them back a year later when funding conditions get better.

Legally designated foundations don't have charity's requirement for regular small donations, so they are not as susceptible to the whims of a large group of people. Instead of regular small donations, foundations instead need to convince a few donors to give a lot of money. Big checks always come with strings attached. While the 'check size' might be higher, it's often distributed over time, so an unhappy donor can pull out, leaving the organization in the lurch. These donors also get lower tax write-offs, which can give a feeling of more ownership over the organization and thus more control over its direction. So while the constraints on foundations are different than charities, they still need to keep donors happy, which could push the organization towards shorter term, sexier

projects that align with philanthropists' priorities.

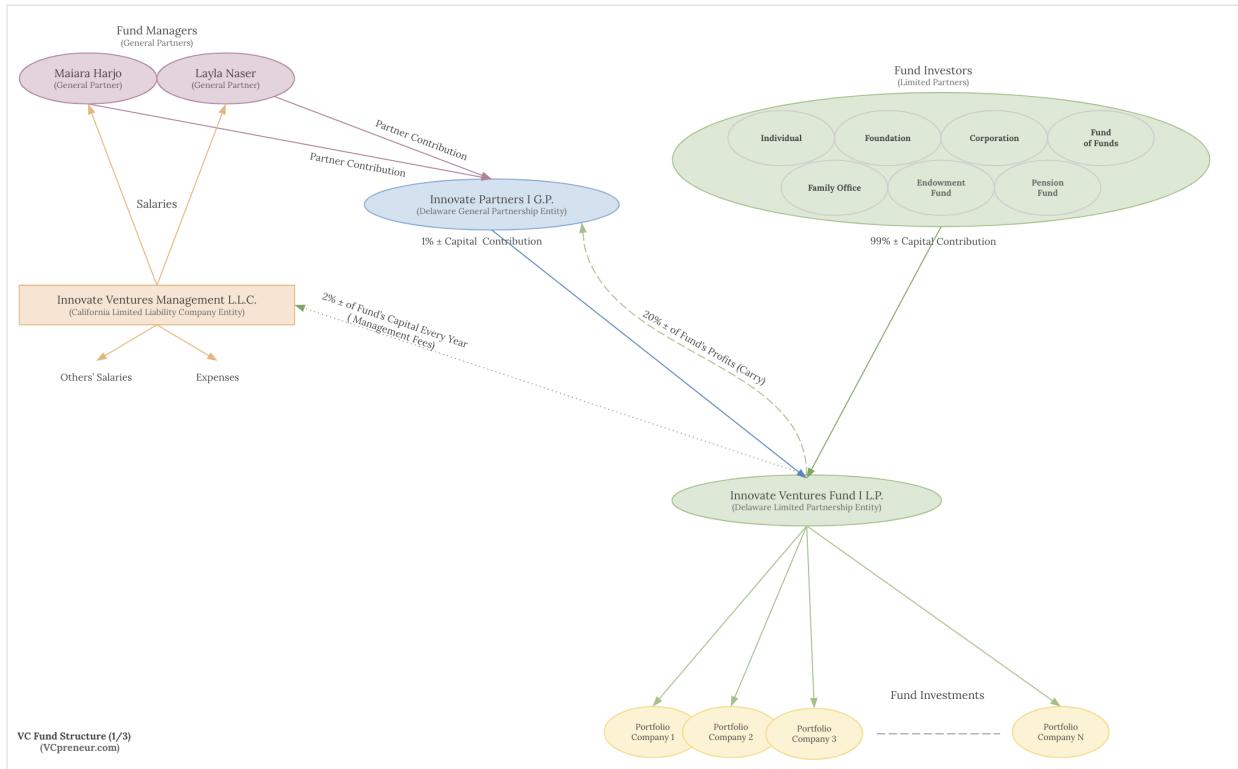
Admittedly, I've outlined worst-case scenarios for how incentives can play out in both charities and foundations. However, people always spend money for a purpose. In the case of philanthropy, that purpose is a complex, fragile 'product' (as opposed to more simple products like eggplants or index funds). Taste in projects is more flighty than the desire for return on capital. People usually want closer involvement when the money is buying a more ill-defined good than "more money."

It's easy for a philanthropy to take actions that do not deliver on donors' expectations. In other words, it's easy for philanthropies to become misaligned with their money factories. This easy misalignment makes it dicey to rely only on donations for long-term projects. Unfortunately, the nature of research means that it both needs stable, long-term funding *and* often fails to live up to expectations. In short, structural constraints make the traditional nonprofit structure a shaky foundation for an organization that requires a significant chunk of capital and produces uncertain results on a long timescale.

There are no first principles for legal structures

Disclaimer: I am not a lawyer.

You'll notice that most organization's legal structures don't stray too far from a few basic models. New companies tend to be vanilla C-Corps if they intend to sell equity and LLCs if not (and their equivalents outside the US.) Organizations with charitable intent tend to be 501(c)(3) designated charities or foundations. VC firms tend to be LLCs with a few members forming a general partnership. This general partnership then creates closed-end funds that are limited partnerships with other investors.



The structure of a typical VC firm¹¹⁰

At the end of the day legal structures are “just” contracts that determine how money flows and how decisions are made both within and between one or more taxable entities. We generally think of organizations in terms of these tax designations: LLC, C-Corp, 501(c)(3), etc.¹¹¹ Our custom of using tax designation as one of the primary descriptions of an organization hints at the fact that taxes matter a lot! Tax designations can place requirements on organizational structure like the need for a board of directors and determine both what happens to money that comes in (eg. C-corps and their shareholders are both taxed, only the shareholders of an LLC are taxed) and what an organization is allowed to spend money on (eg. 501(c)(3)s can only spend money related to their official mission).

¹¹⁰ From VC Funds 101: Understanding Venture Fund Structures, Team Compensation, Fund Metrics and Reporting I by Ahmad Takatkah | VCpreneur

¹¹¹ Less-used tax designations vary both over space and time like S-corps or Benefit Corps. For example, as of late 2020, benefit corporations are authorized in 35/50 US states. They exist in some states but not in other states and are legislated into existence and sometimes out of existence.

These contracts have a big impact on an organization's output. That impact makes it important to find the appropriate legal structure. Moreover, legal structures have more flexibility than most people assume.

People tend to implicitly treat legal structures like elements and particle physics¹¹² — a set of clear options in the same way that Hydrogen, Helium, and Carbon are composed of electrons, protons, and neutrons. That's not the case at all. You can actually write whatever contract all the involved parties are willing to agree to. The question is what will happen afterwards — it could be totally fine or the government could sue you out of existence and send you to jail. It seems like given a set of laws, you should be able to tell what is in-bounds and what is out-of-bounds. In reality, it's extremely hard for lawyers to know a-priori whether a legal structure is legit besides trying it and seeing whether you get sued. There are no first principles that let you derive legitimate legal structures.¹¹³ Composite materials are a good analogy: beyond well-studied or simple examples, it's extremely hard to predict how truly novel materials will perform.

This illegibility is an underlying reason why people tend to stick with legal structures that have already been safely demonstrated. (The same is true with materials!) Limited partnerships are used for almost every small investment firm not because it was a carefully thought-out structure but because it had precedent in the whaling industry where it had basically been made up by common law. Delaware C-corps are everywhere because in part because they have so much case law behind them that the lines of acceptability are much clearer than any other option.

Sticking to well-known structures creates several reinforcing feedback loops.

¹¹² That is, particle physics before subatomic particles, now it's just a mess all the way down.

¹¹³ The property that you don't actually know where the line is until you cross it is true in law more broadly. Some legal scholars in common-law countries like the US consider it a good thing when laws are challenged because more rulings give a better sense of where the line is.

Most lawyers are hesitant to even suggest a non-standard structure because they don't have the experience to intuit the affordances of a new structure. That leads to even fewer lawyers with that experience. People are (rationally!) hesitant to fund non-standard structures. It's generally not worth the work to avoid the risk of lawsuit, surprisingly high taxes, or other surprises. This pressure from backers to stick to standard structures creates an adverse selection effect: the things people are interested in funding are nudged towards traditional structures, leaving only the things nobody wants to fund with weird structures and making it a good guess that most organizations using non-standard structures are sketchy. The association between unconventional legal structures and sketchy organizations wears the habit grooves even deeper, increasing hesitation to do something different and creating more examples of organizations that have succeeded despite a-few-sizes-fit-all legal structures. This survivorship bias ignores all the counterfactual organizations that failed or never even started because of the restrictions they faced.¹¹⁴

This all might seem like cute esoterica, but the combined flexibility and illegibility of legal structures has several significant consequences. Legal structures are intimately tied to the question of how money works in an organization, the actions that an organization can take, and the incentives they can deploy. Don't we want more organizations that can use philanthropic dollars to produce public goods, but then are able to capture some of the value they create in order to produce more things? Or perhaps organizations that look like a company, but instead of implicitly lasting forever, they have finite lifetime like a closed-end fund?

OpenAI's capped-profit/nonprofit structure¹¹⁵ is a great example of an organization playing with legal structures to match their own situation and shape incentives in ways that wouldn't otherwise be possible. They need to be

¹¹⁴ Admittedly, the number of organizations constrained out of existence by people being unwilling to play with legal structures might be zero. But it also might be huge. Counterfactuals are hard!

¹¹⁵ See <https://openai.com/blog/openai-lp/>

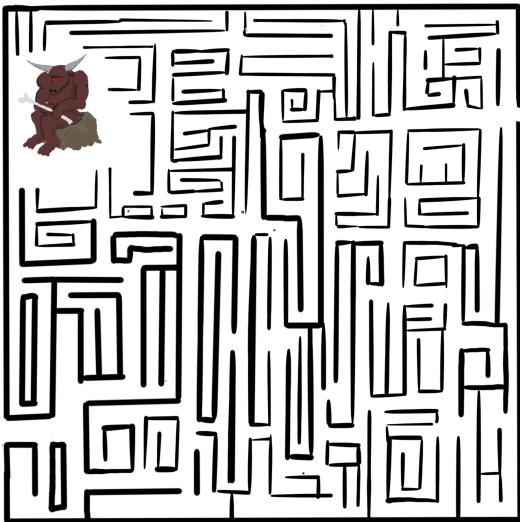
able to pay AI researchers top dollar, while at the same time their nominal mission is to keep powerful AI technology from being captured by organizations whose incentives are to use it to maximize shareholder value.¹¹⁶

At the same time, doing things differently for the sake of doing things differently is rarely a good idea. It's well-worth paying attention to Chesterton's Fence. Most of the time it is easier to use standard legal structures, especially considering the cost of lawyers who understand contract and tax law well enough to play with soft constraints. Old legal structures rarely provide huge barriers to an organization's ability to achieve its goals, just inconveniences. It's a situation where continuous changes can lead to discrete differences — will the little inconveniences from standard structures eventually cause an organization's death (or derailment) by a thousand cuts?

For a DARPA-riff, I would argue the answer is yes.

¹¹⁶ One could argue how much of this is just talk, but I default to trust and it helps argue my point.

Riffing on DARPA



While many people can inspect the outside of the maze and the stories are pretty clear on the relevant features of the minotaur (bull head, human body, kills (and probably eats) people, superhuman strength, etc.) the layout of the labyrinth could conceivably be in any configuration that conforms to the exterior walls.

When faced with a situation where you cannot reason from first principles and relatively long iteration loops there are limited options for thinking rigorously about it. Neither scientific inquiry nor engineering design can function so we have to look at historical narrative.

Imagine we're in some parallel universe where Daedalus was a prolific labyrinth builder and while nobody had ever gone through the modern Labyrinth we've described, there were others who had successfully navigated other labyrinths. The best strategy is probably to learn as much as we can from them and try to adopt their tactics to our own situation. In our case, there is an entire pantheon we could dig into¹¹⁷ but the two mythical organizations that stand out the farthest are Bell Labs and DARPA. Of these two, I would argue that DARPA's idea maze looks much more like the one we face today.

¹¹⁷ There are obviously many organizations with structures worth learning from — a deep discussions of what to learn from each of them is important but beyond our current scope.

Why try to emulate DARPA and not Bell Labs

Bell labs was a product of its time and a set of unique conditions that no longer exist today. DARPA, on the other hand, seems able to work in today's world just fine. Its biggest hamper is perhaps the constraints that come along with being part of the government.

I suspect there is more room to play with the DARPA Model because the 'search space' for replicating Bell Labs is both more restricted and has been heavily explored. Bell Labs didn't have a unique model. While they definitely have the best narrative, there were several other great industrial labs. Many organizations have tried to create "the next Bell Labs." Those that have tried it as a part of an existing company without the right external conditions were crushed under the constraints of modern corporate R&D and those that tried to do it as a separate entity did not address the the trifecta of conditions required by healthy industrial labs. Arguably, we still do have great industrial labs today, just not in atom-based disciplines.

In contrast to Bell Labs clones, extra-governmental riffs on the DARPA model are rare. Many organizations give lip service to the DARPA model but don't resemble it at all. It's easy to set out to build a DARPA but end up building a Skunkworks. Of the organizations that do seem like earnest riffs, [Actuate](#) (solutions R&D for social good) and [Wellcome Leap](#) (solutions R&D for health) have only started in earnest in the past year so it's hard to say much about them. [Ink and Switch](#) seems to be successfully riffing off of the DARPA model in the software world. There is a lot of organizational whitespace still to explore.

At the same time, this apparent whitespace could be the same as shouting "the real non-governmental DARPA model has never been tried!" into a graveyard of failed attempts. Maybe there have been organizations riffing on DARPA that died quietly or didn't explicitly draw the connection between their model and DARPA's. Another real possibility is that the DARPA model just doesn't work

outside the context of the government. However, the evidence suggest that a private DARPA-riff is not impossible, and may in fact be possible. After reading this section, hopefully you'll agree.

In summary, it seems possible to replicate DARPA in today's world but it doesn't seem possible to purposefully recreate the conditions that enabled Bell Labs to be what it was.

There are DARPA-style ideas that DARPA doesn't pursue

A key reason why it would be impossible to riff on DARPA is because actual DARPA is already tackling all the parts of the adjacent possible that fit within its structural constraints. All DARPA programs need to tie into US defense priorities. However, almost any piece of technology can be tied into defense priorities somehow. For example: human-computer interfaces are for command and control, medicine is for helping wounded soldiers and disabled veterans, improved learning techniques can help train military personnel, new materials and energy-generation technology can be deployed on the battlefield and make our military more energy-secure, and fundamental physics could enable warp-drive-powered battlecruisers!

Therefore, it's quite reasonable to conclude that DARPA would already have tried any potentially impactful DARPA-style program. If this were true, trying to run similar programs with less resources and clout would be idiotic.

Luckily, evidence suggests this isn't the case. It's perhaps a cheap analogy, but assuming that DARPA will undertake every potentially-awesome technology program is the same line of thinking as assuming that Google will go after any potentially-valuable software product.

One piece of evidence is simply testimony from former PMs that they had program ideas rejected because they were insufficiently applicable in a defense

setting. It would be lovely to have a list of “good ideas DARPA rejected”, but DARPA is still a part of the US DoD so it’s hard to just ask about those proposals directly.¹¹⁸

Another piece of evidence are stories of how many of the programs eked into existence by the skin of their teeth. Their stories often involve a dedicated PM doggedly arguing for them through several rejections or impending budget cuts¹¹⁹. It *could* be that DARPA’s selection process is actually amazing at supporting almost every eventually-viable idea but puts them all through the wringer first. This seems unlikely. Instead, if many of the DARPA programs, especially those that are especially exciting outside of the military, barely made it over a line, there are as many or more that didn’t make it over that line.

While it famously has fewer constraints than other government agencies, DARPA is still part of the US government bureaucracy. As such, DARPA still has a number of constraints that could rule out otherwise viable programs. DARPA doesn’t do any research in house, so it cannot pursue programs that would require spinning up an entirely new organization. It still needs to use government procurement, hiring, and salary practices, all of which could sabotage potential programs.

At the end of the day though, the only hard evidence to support the claim that there are DARPA-style ideas that DARPA doesn’t pursue is to successfully design and run them!

¹¹⁸ Former PMs are still (understandably) hesitant to talk about could-have-been program ideas.

¹¹⁹ See for example the history of DARPA’s drone program in The DARPA Model for Transformative Technologies: Perspectives on the U.S. Defense Advanced Research Projects Agency

Can the DARPA model exist outside of the government

In addition to establishing the fact that it *is* possible to riff on the DARPA model it's also important to ask whether that riff can exist outside of a government. Is it only possible to riff on the DARPA model successfully with organizations like [IARPA](#) or the recently-announced British ARPA, ARIA?¹²⁰

There is no good way to satisfactorily answer the question, but we may be able to make more headway on the isomorphic question:

What are the reasons that something could *only* exist in the context of a government?

Put differently, there are 'hard' and 'soft' reasons why an organization would need to be part of the government. Hard reasons are reasons that you literally *cannot* get around without a government. Soft reasons are places where it seems *really hard* without a government but possible. If the only reasons we can find are soft, we are forced to conclude that a DARPA-riff can exist outside of a government.

There are two hard reasons that an innovation organization needs to be part of a government. The first is that anything whose core work involves controlled technology like things that can directly be turned into nuclear- or bio-weapons. In these cases, if you aren't part of a government, some government will, with varying degrees of force, make sure that you don't exist at all.¹²¹ The second hard reason an organization would need to be part of a government is when it requires more capital than could possibly be aggregated without the power of a nation-state. The original moon landings are good examples of work that could not be accomplished without government funding¹²² — NASA's budget was 4%

¹²⁰ See A new UK research funding agency

¹²¹ And even then it depends on which government you are part of.

¹²² I want to caveat that the moon landings could not have been accomplished without government funding in the 60s. Whether something requires government-level funding

of the US federal budget in 1965 and 1966.

All of the other reasons feel soft. There are probably more soft reasons why an innovation organization would want to be part of the government than are useful to enumerate. I'll list a few.

- The work involves doing things that impinge on people's rights (privacy, property, etc.)
- The work is expensive without much possibility for capturing a return.
- The work creates public goods or outputs whose value can only be captured on the scale of a nation-state.
- The weight of government authority could significantly advance the work.
- The organization needs a recurring budget that it can't find as part of a corporation.
- The organization needs to utilize equipment and resources owned by the government like national labs.

An organization riffing on the DARPA model absolutely is subject to several (all?) of the soft reasons. However, I would argue that it isn't subject to either of the hard reasons. We have actual DARPA to do weapons research and the numbers necessary to run actual DARPA are large (\$3.6B in FY2020), but not impossible to achieve without government funds.

While being part of a government could certainly relax some constraints standing in the way of success it *imposes* many other constraints. Successfully riffing on DARPA might be extremely hard without being a government entity, but riffing on DARPA will be extremely hard *as a government entity* — just in a different way. Two actions of former DARPA directors Regina Dugan and Arati

is not an inherent property of that thing but instead depends on the nature of the work at the time, the relative wealth of the society, and the concentration of that wealth. In the 1960s, NASA needed to do a ton of new things, the US was not as wealthy as it is now, and that wealth was less concentrated (fewer billionaires).

Prabhakar seem to back up this conclusion — both of them are working on non-governmental ARPA riffs. Dugan is the head of [Wellcome Leap](#) and Prabhakar started [Actuate](#). I agree with their tacit conclusion and would argue that in addition to being viable without government support, 21st century riffs on DARPA *should* be private.

21st century DARPA-riffs should be private

On the back of evidence that it's possible to build a DARPA-riff outside of a government, I want to make a more aggressive claim: 21st century DARPA riffs not only can, but *should* be voluntary organizations. There are two major reasons new DARPA-riffs should be private. First, modern bureaucracies introduce an intractable tension between accountability and the opacity a high-variance organization needs to succeed. Second, government-independence gives an organization a fighting chance to do better than (actual) DARPA in some areas by taking advantage of the ability to do things that a government organization cannot.

If you ask most people “do you want the government to be accountable for the money it spends?” They would probably answer “of course!” In a liberal democracy the government is nominally accountable to The People; as much as possible, we demand transparency and oversight over how our money is spent. Government transparency is good but it also stands in direct opposition to the fact that arguably opacity is important to DARPA’s outlier success¹²³. If we are painfully honest, the actions that lead to great results often *a priori* look like bad ideas or offend our sense of fairness. It is completely reasonable to demand that an elected government be fair. However, making a game more fair reduces its variance and rules out many activities; the way ARPA operated in the ‘60s would *never* be allowed now. These two conflicting but correct outlooks put any government organization trying to produce high-variance results between a rock and a hard place. The way to slice through this Gordian knot is by not forcing the

¹²³ See Why Does DARPA Work?

government into the position in the first place.

DARPA isn't perfect.¹²⁴ While some of the downsides, like high variance outcomes, are structurally tied to the model itself, others are artifacts of being a government organization. As a government organization, DARPA needs to have an air of impartiality — they must issue open calls for most programs, go through a rigorous process to demonstrate that they considered all options before making a decision, work with legitimate institutions, show are paying no more or less than fair market price, the list goes on. On paper these are all good things, but they also create a slower more constrained process that potentially leaves opportunities for a private organization. A private organization could potentially tap talent that is hard for a government organization to attract because it has much more flexibility to play with salaries, equity, and other incentives. Hopefully this list (that could be expanded) is convincing evidence that there are potential improvements to the DARPA model that are out of reach for government organizations.

¹²⁴ See the Room for Improvement section of Why Does DARPA Work?

A DARPA-riff both can and must do multiple experiments at once

Running multiple organizational experiments flies in the face of common wisdom that you should only change one thing at a time.¹²⁵ This is absolutely true for building a successful business. However, people have tried millions of fairly open¹²⁶ experiments with organizational practices over many decades in everything from human resources to management to employee compensation. As a result, while you might not like how an organization does HR or design, basic organizational structure and best business practices are pretty effective. Research organizations have simultaneously had many fewer experiments (just compare the number of companies that have been started to the number of university labs that have been started), less selective pressure, and looser feedback loops.

Concretely, you're in a bind if you simultaneously believe it is worthwhile and possible to riff on DARPA and that you should only change one thing at a time. You would be limited to creating new government organizations like ARPA-E and IARPA. If you buy the argument that it's possible to riff on the DARPA model outside of government that single shift forces you to rethink multiple things (legal structures and funding for a start) whether you like it or not. At that point, clinging to existing structures could become more of a liability than wisdom because the assumptions that made those structures good ideas in the first place no longer hold.

Setting up a new colony in Massachusetts is very different than on Mars. Environmentally, Massachusetts is pretty similar to England so it makes sense to - sure you might want to use some different crops to deal with the weather but you still want to use roughly the same tools, farm the same way, and build houses the same way. Even the extreme religious stuff that was the whole point

¹²⁵ I like to think of this like how polynomial approximations are valid in a local area around an equilibrium.

¹²⁶ There are few things more fun to talk about than how messed up your old company was.

of the colony eventually reverted to the mean. Mars, on the other hand, is unlike *anywhere* on earth (except maybe Antarctica). If you tried to live there in anything resembling the way you did before, you're dead.

So a DARPA-riff needs do some institutional experiments to survive at all. On top of that, There are a number of experiments that seem like they can only be done in the context of an organization. Since the goal of a DARPA-riff is not just to enable more paradigm-shifting technology but to try to demonstrate a new institutional model, guinea pigging should be part of its role. A laboratory for experiments in 21st century research management, if you will.

This *isn't* to advocate throwing everything away just because it smells like the past. Chesterton's Fence is important! Instead, it's important to learn from the those who have come before — hopefully the valuable lessons from Bell Labs and DARPA have convinced you of that. But in a strongly disequilibrium state, you need to evaluate each component on its own merit and perhaps grab components from disparate parts of history or the organizational landscape.

Pulling off this balance will be like sailing between Scylla and Charybdis; extraneous experiment on one side and the same old incentives leading to the same old outcomes on the other. With this gauntlet before us, I want to note that there is a whole ontology of both experiments and structural-components-that-shouldn't-be-poked. There are some parts of the DARPA model that if changed, would make it hard to even call an organization a 'DARPA-riff' in the first place. On top of these immutable considerations, there are some experiments that *must* happen figuring out how money works without a DoD budget is one of the most glaring ones. Other experiments are more 'optional' — they might make the difference between success and failure or they might be a distraction.

A DARPA-riff needs to be obsessed with program managers

The relationship between the DARPA Model and empowered program managers

is like the relationship between a hydra and its many heads. The heads are part of the hydra, its business end, and one of its defining characteristics.

[Why Does DARPA Work?](#) covers DARPA PMs extensively, but the key points are worth reiterating both as a refresher and to emphasize that empowered PMs are the core thing *one should not mess with*. “What about managing programs through a committee?” NO. “What if performers just submitted grants and coordinated among themselves” NO. “What about a more rigorous approval process to make sure money isn’t wasted?” NO. “What if people could be career program managers?” NO. You get the point.

The importance of program managers means that anybody attempting to replicate DARPA’s outlier results needs to be utterly obsessed with bringing the right program managers on board. Getting the right program managers will require understanding both who those people are and how to convince them to join.

Program managers can’t just be general high quality people but need to exhibit specific traits: curiosity; low ego; good at bringing people together; good at communication; and have high doing to talking ratios.

A DARPA-riff will need to do some hard work to convince good program managers to join. A new organization can’t depend on accumulated prestige and a private organization can’t take advantage of patriotic instincts the way that a government organization can. So you’ll need to be creative! One speculative approach for recruiting PMs is to embrace people who have expertise but are not traditionally credentialed. This pattern is not uncommon in software-adjacent disciplines like deep learning or cryptography where people can independently build mastery and then do real cutting-edge work without going through traditional institutions. Outside of software, uncredentialed research expertise is rarer; developing it requires more specialized equipment and tacit knowledge that lives primarily in traditional institutions. However, uncredentialed still exist outside of software and if predictions about the democratization of disciplines like biology play out, they may become easier to find over time. Developing taste

in uncredentialed experts might yield fruit!

A DARPA-riff has career affordances that you can't find elsewhere which will hopefully appeal to a subset of folks who would be excellent PMs. There aren't many opportunities to work with extreme autonomy on getting a technology out into the world without needing to tune the work to either produce papers or hockey-stick growth numbers. I suspect that there are a few extremely competent people out there (it doesn't take that many) who are seething to shift technological paradigms and are willing to trade off salary, prestige, and stability for the ability to make those shifts happen. Of course, often that sort of pitch leads to bitter disappointment, so smart people are rightly skeptical of it! The burden of proof is then on the organization to become a place that does deliver on those promises.

More than any specific ideas though, an DARPA-riff will need to constantly ask itself "how do we become a place that PMs want to work?"

People who have done research should be in charge of research (and more generally, people who have done a thing should be in charge of that thing)

Once hired, PMs have significant agency to make programs succeed or fail. This power should make anybody building a DARPA-riff take very seriously the question "what makes people good at being in charge of research and how do you find those people?" This question is also important for anybody who cares about innovation systems more generally — too-involved-funders, MBAs running the details of research programs, and other control-project mismatches are all-too-common.

"Being in charge of a thing" is a nebulous concept. However, at its core is the ability to directly control the fate of that thing; some combination of deciding whether a project or person gets approval funding, how well a project is doing,

and coming up with ideas for what should be done in the first place.

In research especially, there are very few metrics you can use to distinguish between a good idea and a bad idea or how a project is progressing once it's started. Even when metrics suggest that "everything is working great!" it is easy for researchers to fool themselves, optimize for the wrong thing, or overlook a core problem. As a result, the decisions that someone in charge of research needs to make demand a level of understanding that can peel back the metrics and poke at what's going on underneath.

The 'understanding' that allows you to see past metrics has a lot to do with intuiting the affordances of that thing. That is, you have built up a rich enough tapestry of patterns that you can subconsciously integrate many explicit and implicit inputs and have a sense of the effect of different interventions into the system. This process, almost by definition, is hard to abstract — if we could deabstract it well, you could put it in a metric!

To illustrate, I'll use an example that has nothing to do with research: gym managers. You can always tell when the manager on duty at a gym is a weightlifter themselves by how they react to people deadlifting in socks. Deadlifting in socks is a common practice; nominally because it enabled you to 'feel' the floor as well as possible and balance while lifting heavy weight (and also because some famous weightlifters recommended it many years ago). Rules that everybody must wear shoes in the gym are also common practice; shoes can mitigate the damage if you drop something heavy on your foot and some people would find it gross if everybody was walking around without shoes.¹²⁷ If a manager hasn't done much lifting, they'll immediately stop someone deadlifting in socks — it's against the rules! (Note here that metrics are nothing but rules about what's good and bad.) If a manager has lifting experience, they'll generally let someone deadlift in socks *if* they look like they know what they're doing and are lifting enough weight to warrant it. Not only will they probably not hurt themselves, but shoes won't stop damage above a certain weight anyway. Note

¹²⁷ These people should not visit climbing gyms.

even in this simple example how quickly we run into illegibility! What does it mean for someone to “look like they know what they’re doing?” It’s not just a matter of how big their muscles are or whether they use chalk — it’s how they put the weight on the bar, how they hold themselves, the ‘intention’ in their eyes; ie. things that are completely inadmissible as metrics. Research is immensely more complex than a gym!



This person looks like they know what they’re doing.

As uncomfortable as it is to say in our metrics-driven world, both good gym managers and research managers depend on intuition. This section should more accurately be titled “people who have an intuition for research should be in charge of research.” However, it’s extremely hard to get a sense of someone’s intuition for a thing — even if you yourself have intuition for thing. So in a delightful meta-situation, we need to find more legible heuristics for ‘intuition.’

“Having done a thing” is a reasonable heuristic for “has intuition about this thing.” It’s of course not always true; people who have done a thing do not always have intuition for it and people who have not done the thing can sometimes build intuition it! It helps if they have a track record that could only be accomplished by having an intuition for the thing or extreme luck. Of course,

it requires intuition to distill what pieces of a track record can be accomplished without intuition. Intuition turtles all the way down.

The “gotcha!” comes in deciding what counts as ‘having done the thing.’ Does someone who worked in a biology lab doing experiments on rat brains count as someone who has ‘done’ biology? Should they be in charge of a molecular genetics project? A DNA-origami project? This is where you need to defer back to the fact that experience is just a heuristic and attempt to tease out how much intuition is shared between the past experience and the future work. The trap to avoid here is that sometimes someone will overspecialize and end up being *less* qualified to be in charge of a thing in an area that’s normally quite close to their experience. For example, someone who is a tenured professor in biology working on lipid bilayers might be terrible at being in charge of a CRISPR-based program, while someone with relatively less experience would be good at it. The trap *there* is to take that logic to the extreme and declare “experience doesn’t matter!” If you take anything away from this section, I hope it’s that experience *does* matter but that it isn’t the only thing that matters.

Of course, at some point someone needs to be in charge of several things and they can’t have done everything. My hunch is that the trick to navigating this unavoidable tension is ‘attenuated delegation.’ That is, the less intuition someone has, the more heavily they should delegate or conversely, the less they should be involved. You could roughly divide involvement into three levels: binary go/no-go decisions, progress measurement and course correction, and directly helping in the process. Each of these requires quite a bit more intuition than the one before it.

Anybody attempting to emulate DARPA needs to be willing to sit with a lot of discomfort for a long time

The results of solutions R&D take a long time to impact the world. On top of that,

the attribution is usually mixed because of how many entities often need to touch a technology between where a DARPA-riff will work on it and when it enters regular use. I desperately want to burn this into your brain because I suspect a private DARPA's most likely failure mode is the frustration of some combination of people working on it or supporting it, leading to result-killing timescale compression or abandoning the organization all together.



The Mother of All Demos happened in 1968 — six years after J.C.R. Licklider joined DARPA. Personal computing technology took another five years to go from this demo to prototype system in the Xerox Alto and (depending on how you count) another ten years after that to become a commercial product in the Apple Lisa. That's 21 years between the start of the ARPA program and a product on the shelf that people can buy. Similarly, The first DARPA Grand Challenge that arguably kicked off modern autonomous driving happened in 2004 (and was arguably a failure — none of the cars could complete the course) and of this writing in 2021, autonomous vehicles are still not mainstream. That's 17 years and counting.

People in the startup world often pride themselves on their timescales — ten years is the rough average from first funding to IPO. The timescales we're looking at here are twice that or more, not to an IPO, but to a customer-facing product (if the work even ends up as a product!) that's possibly attributed to someone else entirely. It's like running a relay marathon. Of course, some scientists are rolling their eyes — “that's nothing” they say, tuning the experiment they've been working on for the past three decades.

Adding to the uncertainty is the fact that a big part of solutions R&D is to get the ball rolling, not get it to the goal. This role is important but muddles attribution. You see this historically in how DARPA seeded some of the best CS departments in the country; DARPA doesn't get credit for everything that has come out of the University of Utah CS department ¹²⁸(nor arguably should it!) DARPA money often plays a role of derisking research to the point that the NSF is willing to fund it. Almost by design there will be arguments about who deserves credit for things a DARPA-riff helped with.

On top of excruciating timelines and attribution uncertainty, riffing on DARPA will entail working on *weird shit*. You've heard of the successes, but DARPA

¹²⁸ The University of Utah CS department is notorious for early computer graphics work like the Utah Teapot. Pixar Founder Ed Catmull is a direct descendent of this tradition.

regularly funds wacky things that go nowhere. Things that make you think “they’re wasting my tax dollars on *that*?” Warp drives. Remote control insects. Creating fusion through sound waves. Super-soldier serum. Though it should be noted that DARPA had nothing to do with telepathic goat murder research. Over time, DARPA directors have needed to fight to keep DARPA from being toned down and turned into a normal R&D org.

A big part of the unavoidable discomfort for supporters of a good DARPA-riff is that any of these attributes can also be used as an excuse for poor performance. “We’re not showing clear results because we need more time to resolve uncertainty” can be true or false and even the person saying it might not know the difference. The line between a program that just takes time and is going nowhere is often illegible to everybody but the people working on it. Ditto for attribution and weirdness. Here lies yet another Scylla and Charybdis that we must navigate; there is no compass here besides trust.

It is important for big things to start small

As far as I can tell, every big thing that really sticks in people’s heads started at bare minimum an order of magnitude smaller than where it ended up. This goes for companies, organizations, projects, movements, you name it.

By contrast it seems that whenever someone starts something huge — a massive fund, a company that is hyped from day one, a movement where they declare “we are starting a movement to change the world!”, a new field where they declare “we are creating a new field!” — it ultimately disappoints.

There are several practical reasons for why big things tend to succeed more often if they start small. Three of the biggest ones (which are not completely independent of each other) are expectations, momentum, and trust.

When something starts small, it has the ability to exceed expectations. Exceeding expectations sticks in people’s heads more than meeting them. Exceeding even small expectations is memorable! By contrast, when something starts big, it has

massive expectations from the beginning so it is hard to exceed them.

Small things have less momentum, which in turn allows you to make many quick, micro-adjustments to iron out problems and move in a better direction. Less momentum lowers the cost of changing things. It's the difference between being in a kayak and a cruise ship moving at the same speed. If you know you're going in the right direction, you want to be in the cruise ship, but almost nobody gets everything right from day one. You want to turn the low-momentum kayak into the cruise ship only once you've made several course corrections.

The momentum-related costs of changing direction come in several forms. Direction-changing costs could appear as coordination-related transaction costs between people: the more people who are involved in something, the more coordination costs you accrue when you try to change anything. The costs could be in terms of reputation: the more eyes and expectations on what you're doing, the bigger the confusion and or reputation damage you cause when you choose to do something different. The costs could be literal capital costs: if you buy an expensive piece of hardware that ends up being useless or a large office in a place that ends up being the incorrect place.

Starting small enables you to build up trust. In large part, building trust is a process of verifying that someone or something is both competent and reliable. To some extent it's just about "getting in the reps." The same quick, small actions you use to course correct can (if done competently) create many trust-verify loops. A history of reliable and competent small actions provide a strong foundation for bigger leaps of faith. On the other hand, asking people to take a big leap of faith up front often leads to second-guessing or sandbagging, which in turn can bring even the largest organization to its knees.

In the context of a private DARPA, this is warning about the dangers of coming out of the gate yelling "we're going to change the world!" And raising a massive war-chest commensurate with that statement without first working out kinks and building trust. It's *not* an injunction against massive ambition! Soberingly, the admonition to start small creates a tension in conjunction with the

‘portfolio’ nature of the DARPA model, which we’ll address later.

The boon and bane of institutional missions

On the note of ambitions, it’s worth briefly talking about institutional missions.

Institutional missions are a tricky thing. On the one hand, they can be an inspirational rallying point and a powerful filter for actions you should and should not take. On the other hand they can be a distracting pile of words that obscures reality and enables people to feel good about themselves without earning it.

It’s unclear whether people focus too much on missions or too little. Most mission statements are bullshit on which people fixate at the expense of putting their heads down and letting actions speak for themselves. This phenomena encourages (at least in me) a counter-reaction to eschew thinking and talking about mission entirely — “let’s earn the luxury of thinking about mission through hard work first.” I’m still sympathetic to this view, but at the same time spending some cycles on missions is probably worthwhile. People use institutional missions as a proxies for what that organization actually does, so without a mission it will be hard to succinctly get across a useful mental model of an organization. It’s especially hard to build that model if you’re doing something new with long feedback loops like, I don’t know, building a private DARPA. All institutions have implicit missions, so there *is* a mission whether you make it explicit or not.

A good mission can also be a powerful filter both on actions and on people. You should be able to ask of any action or project “does this fit within the mission?” And be able to come up with a concrete answer that can plausibly be “no.” Creating a mission that sounds good to everybody is seductive — we all want to be loved and be lovely. As a result, most missions are so nebulous as to arguably cover anything. They don’t enable you to rule things *out*. Any given mission shouldn’t appeal to everybody. Institutions need to differentiate between individuals inside and outside the institution and missions that don’t sound

good to everybody can help with that.

SpaceX and Bell Labs' missions are good examples of broad but discriminating missions. Bell Labs' mission was to "Support the Bell System with Science and Engineering Advances." SpaceX's mission is to "Make humanity an interplanetary species." Both of them cover a *massive* range of activity, but at the same time there are many activities that do not fit under their umbrellas. Both of them will appeal to some people and completely turn off other people.

Frontier technology has bespoke sales channels

Outside of software and therapeutics, there are no good default paths for new technologies to get into the hands of end-users.¹²⁹ The less it looks like anything that came before it, the fewer precedents there are for how to buy or sell it! Figuring out how to sell a technology¹³⁰ can be a comparable challenge to developing it in the first place. It's a challenge that's worth worrying about even if your primary goal is not to make money because someone needs to buy manufactured technology eventually in a monetized economy. Questions about sales channels introduce a tension into technology creation — things you can do to make a technology more useful to end users (for example, which performance metrics you focus on) but too much focus on end use can push the technology into a premature local optimum.

By contrast, software and therapeutics have well-established sales channels. As the two (glaring) exceptions to the rule, they are illustrative and perhaps hint at how to improve sales channels for other technologies. People have sold software for long enough and in enough volume that there are well-established strategies

¹²⁹ There's not a great word for these people. "End-users" sounds incredibly clinical and alternatives like "consumers" invokes large numbers of unsophisticated people.

¹³⁰ (Who/which member of an organization is actually spending the money? Who do you need to convince that it's a good idea? How does the customer know it works? What does the contract look like? Do you need to go through a third party? These are of course questions for any product (and invite innovation!) but more established types of products have more standard answers.

for diffusing software: ‘land and expand’, enterprise sales, open source while selling managed instances, etc. Individual consumers and companies have reasonable mental models about how to acquire and test new software. This isn’t to say that everything has been tried, that software sales is easy, or that software is all the same. However, there are fewer unknown unknowns on both sides of a software transaction than, say, a new sort of hydrogen filter for a petroleum processing plant.

The FDA approval process provides a set of extremely legible milestones that allow therapeutic developers and funders to resolve the normal Mexican standoff: “We’ll tell you how much money we’ll give you if you can tell us some clear milestones.” “We’ll be able to tell you some clear milestones if you give us a sense of how much money you’ll give us.” Instead the negotiation becomes “We need \$X to get through Phase 1 - if we do that, will you agree to give us \$Y to get through Phase 2?” “Sure.”¹³¹ This relatively straightforward (but still risky!) path from lab to market depends on the incredibly low market risk for therapeutics that make it through the FDA. Insurance companies pay for the majority of therapeutics and regularly agree to how much they will buy and how much they will pay for a given therapeutic assuming it makes it through the approval process. A big part of the job for more traditional VCs who focus on therapeutics is to lubricate this whole process.

Figuring out how to sell a technology can be as complicated as creating it in the first place: *Who/which member of an organization is actually spending the money? Who do you need to convince that it’s a good idea? How do they know it works? What does the contract look like? Do you need to go through a third party?* All of these questions slither subtle tentacles back to the development process. These tentacles raise many uncomfortable questions — how low does the bill of materials need to be? What does this need to interface with? Should this even be a product? (“build a startup” is in itself a channel design choice). Uncertainty about a technology’s sales channel in turn affects how much support it can in the first place. People are

¹³¹ Obviously there is much more haggling involved but it is a much more straightforward conversation for both sides.

far more willing to support uncertain work like pharmaceutical development that has a clear light at the end of the tunnel.

People creating technology often do not have the bandwidth to give sales channels the attention they require. This limited bandwidth creates another tension — what is the role of the people building technology in diffusing¹³² it? On the one hand, the idea of “build technology and throw it over the wall” has serious flaws — small nudges in development can have large effects on its usability and involved creators can be essential to successful diffusion. On the other hand, imposing diffusion-related questions at the wrong time can be distracting at best and at worse force a technology to a local optimum where it is ‘useful’ but blocked from reaching its full potential.

It’s tempting to punt on worrying about the challenges of diffusing new technology until it’s actually built. The instinct is reasonable but misguided because (un)certainty about diffusion feeds back and hinders or helps technological development. Drugs with billion-dollar development bills can be created only because therapeutics have well-established sales channels. There’s a feedback loop between the availability of funding and what gets built! 80% of VC dollars go towards software and pharma¹³³ which invariably leads to more work in those areas, further establishing sales channels and reducing uncertainty in future software and pharma investments. Conversely, the sentiment “even if it works, you have no idea how you’re going to sell it (or otherwise enable it to have an impact)” makes people hesitant to fund development work or put in the effort to get the technology to work in the first place!

I am not confident about any of the many possible experiments a DARPA-riff can do to address the sales channel dilemma, so I’ll describe them further on in the ‘speculation’ section. However, I am confident that any DARPA-riff that wants to have an impact needs to address the dilemma somehow.

¹³² This word also feels insufficient, but is the commonly-used word for “getting technology to end-users.” I’m going to be a bit sloppy and jump between “selling” and “diffusing.”

¹³³ See the NVCA 2017 Yearbook

Stepping into the shoes of industrial labs

If we've established anything so far, it's that a DARPA-riff will need to fill at least part of the niche once filled by industrial labs. Embracing that role, it's important to ask: which aspects of the niche is a DARPA-riff well-suited to fill? Where will it be weak? What can it structurally do to maximize strengths and minimize weaknesses?

1. Industrial labs enabled work on a general purpose technology before it's specialized.

Working on general-purpose technology will require a DARPA-riff to explicitly and uncomfortably push back against the (accurate) common startup wisdom to "focus on a niche."¹³⁴ At the same time, any technology *does* need to find an initial niche eventually so programs should have some precise hypotheses about an eventual application. Counterintuitively, this tension between considering how a technology will be used and potential local-optimum traps may be one of the barometers for a healthy organization, ensuring that it doesn't lean too far one way or another. Maintaining this unstable balance will require a lot of work to frame goals correctly. It's perhaps a cliche at this point, but the development of the transistor is a good example of walking the general-specific line — it had a precise goal: "find a solid-state replacement for the vacuum tube that we think will be useful in telephone repeaters and switches" but nobody constrained the work to only "make more efficient repeaters and switches."

2. Industrial labs enabled targeted piddling around.

A DARPA-riff can enable targeted piddling around through low-stakes seedling programs, using a single organizational umbrella to quickly spin up and close down programs, and having some number of 'free radicals.'

DARPA uses low-stakes seedling programs to 'acid test' the riskiest part of a

¹³⁴ At least at the program level.

program idea.¹³⁵ There's no reason that a DARPA-riff can't shamelessly copy that strategy directly. The trick is to go into them with the explicit attitude "let's see what will happen." It's important to acknowledge that while everyone would rather seedlings lead to a promising program than be a nothing-burger, reality might have different plans. The way to prevent that preference from morphing into pressure is to explicitly make sure that negative results don't have lasting effects. Specifically, negative results absolutely cannot affect a program manager's ability to run other experiments or affect an external researcher's relationship with the organization as a whole.¹³⁶

It's tempting to ask "why do you need a DARPA-like organization at all? Just run each program separately." An organizational umbrella over programs can turn targeted piddling around into a compounding long-term benefit.¹³⁷ The results of piddling around often come in the form of informative failures or tacit knowledge. It's easy for these results of piddling around to evaporate. If each program was a separate organization, pursuing any serendipitous discovery that didn't help the program move towards its goals would be a distraction. Someone might keep track of those discoveries but there would be relatively little incentive to do so.¹³⁸ An umbrella structure can enable serendipitous potential side-quests discovered during the main body of a program to be fed back to the beginning of the "pipeline" and possibly become its own program.

A DARPA-riff could also enable targeted piddling around by explicitly having

¹³⁵ See the section DARPA uses low-stakes seedling programs to 'acid test' the riskiest part of a program idea from Why Does DARPA Work?

¹³⁶ To paint the counterfactual of explicitly making sure experiments have no negative consequences: One could imagine that after a failed seedling experiment a PM could implicitly or explicitly need to 'redeem' themselves by going above and beyond. Similarly, the organizational attitude towards a performer who ran a failed experiment might implicitly shift towards "well they're good people but a bit incompetent."

¹³⁷ As opposed to something like flossing where people are like "yeah I really should be doing that ... later."

¹³⁸ This is an example of the more general principle that optimizing individual components of a system is often at odds with the system itself.

people whose job is to piddle around! These ‘free radicals’¹³⁹ could be competent individuals who aren’t attached to any specific project. They could help PMs design programs, do experiments, or research areas that nobody in the organization has even thought about. You could imagine this role as having the flavor of a ‘pre-PM <X>-in-residence.’ The role would resemble an internship in the sense that the radical isn’t doing anything mission-critical and could potentially become a PM but without the expectation of that outcome. The role would be unlike an internship because it would involve much more independence and there would be less of a teaching component — people in the role would already be skilled. It also rhymes with the constant theme in Bell Labs narratives where ‘free radicals’ between projects would play a critical role in everything from semiconductor manufacturing techniques to solar cells.¹⁴⁰ However, this sort of role would need to be defended: it’s anathema to efficiency and could easily be a waste of resources!

3. Industrial labs enabled high-collaboration research work among larger and more diverse groups of people than in academia or startups.

A DARPA-riff can be a bridge between different worlds and disciplines. Seedling programs enable PMs to slowly pull together a diverse set of performers by holding regular closed-door meetings (so people can talk about real problems instead of just showing off good results) and making sure different performers know each other.¹⁴¹ J.C.R. Licklider was a master of this — not only having regular meetings for PIs working on different aspects of interactive computing but for the graduate students working on it as well. A DARPA-riff can further enhance these collaborations if it explicitly pulls people together into a single organization as the program progresses.

¹³⁹ In chemistry, free radicals are atoms, molecules, or ions with an unpaired valence electron that makes them extremely reactive.

¹⁴⁰ See The Idea Factory again.

¹⁴¹ See the section A large part of a DARPA program manager’s job is focused network building from Why Does DARPA Work?

Interactions between the program managers could be a high-collaboration between wildly different disciplines. Program managers can and should be drawn from a wide range of professional and technical experience. This diversity, combined with correctly set cultural expectations could go a long way towards filling this aspect of the niche. For example, at Bell Labs "They were not to work with their doors closed. They were not to refuse help to a colleague, regardless of his rank or department, when it might be necessary."¹⁴²

Realistically, a DARPA-riff won't be able to fulfill this role as well as a golden-age industrial lab. Programs that initially revolve around externalized research might make high-collaboration environments challenging and there won't be the same number of 'floating' people. However, a DARPA-riff will fill this part of the niche much better than academia or startups do today.

4. Industrial Labs enabled smooth transitions of technologies between different readiness levels — they cared about both about novelty *and* scale.

This one is tricky. If DARPA's small size is indeed a key to its success, a DARPA-riff can't act like an industrial lab by supporting both a large number of people working on extremely speculative projects and a large manufacturing team.

A scheme where programs are set up so that they can bud off into their own organization could address the problem by *becoming* the organization on the other end of the handoff. As they progress, programs could shift from loose collections of projects to independent organizations whose purpose is to get the technology that they developed out into the world; either by selling it as a startup or helping people adopt it as a nonprofit.

Of course, many technologies are more effective if they're absorbed by existing organizations; this is especially true for new processes.¹⁴³ Another important approach to the handoff problem will be involving organizations that the technology might be handed off to from day one (or before!).

¹⁴² From The Idea Factory again.

¹⁴³ See Fundamental Manufacturing Process Innovation Changes the World.

For a small DARPA-riff, smooth transitions will require long-term relationships¹⁴⁴ with both exploratory researchers and manufacturers. Ideally these relationships could become part of a relatively systematized pipeline¹⁴⁵ reminiscent of Flagship Pioneering with less profit-focus and more discipline diversity. On one end of the pipeline you want people helping you surface and smoothly build programs around promising research. On the other end you want to have trusted connections who can help ‘manufacture’ the output, whatever that entails. How to build this pipeline this is an open question. While it will inevitably involve a lot of legwork, it could be systematized by creating a consortium of partners similar to the MIT media lab or the Santa Fe Institute.

5. Industrial Labs provided a default customer for process improvements and default scale for products.

This part of the niche is strongly coupled to the previous one, and comes with similar possibilities and challenges. Smooth technology transitions at the end of a program basically demands something resembling default customers and scale. A DARPA-riff will never have its own Western Electric. However, a persistent organization with a track record of quality work can get into a situation where every program doesn’t need to start from scratch when it is ready to ‘graduate.’

6. Industrial Labs often provided a precise set of problems and feedback loops about whether solutions actually solved those problems.

This is a part of the niche that a DARPA-riff can absolutely move into, but it will require a good chunk of conscious experimentation. Feedback loops that generate precise problems and filter solutions are not structurally built into a DARPA-riff. At the same time, other prerogatives around graduating programs can serve double-duty. Additionally, like many independent research organizations DARPA-riff could do consulting work to help external

¹⁴⁴ The term ‘Relationships’ is usually a suitcase handle word so I’m going to strive to be as specific as possible here

¹⁴⁵ Ack! So many suitcase words!

organizations implement new tools/ processes. In addition to generating feedback, this sort of consulting could nudge larger organizations towards being customers (see #5 above) and provide an alternative revenue source. As mentioned earlier, a DARPA-riff could also set up a consortia similar to the MIT Media Lab, which would give the organization a default reason to talk to other organizations. A consistent stream of new PMs with diverse set of industry experiences could create another feedback loop.

7. Industrial Labs provided a first-class alternative to academia where people could still participate in the scientific enterprise.

This one is tough. A single DARPA-riff might be able to offer an alternative to academia for some people, some of the time but it won't be able to change the ecosystem. However, could dream *very* big and imagine that an alternative could emerge from an ecosystem of DARPA-riffs and other new types of research organizations.¹⁴⁶ A DARPA-riff's programs will most likely still lean heavily on academic work, especially early in their lives. As a result, it won't do as well at providing an alternative to academia as industrial labs. You could and should also support *other* alternatives to academia by working with extra-academic researchers instead of using academic affiliations as a proxy for competence.

8. Corporate labs enabled continuous work on projects over 6+ year timescales.

One might expect that it would be hard to support long-term projects in a DARPA-riff because PMs with relatively short tours of duty is a core part of the model. However, DARPA has a strong precedent of programs outlasting the tenure of PMs who started them. There's no reason a DARPA-riff couldn't do the same thing. Arguably, PM-PM handoffs can maintain continuity better than many academic labs where most of the work is done by grad students with ~5-6 year tenures with very little incentive for smooth handoffs¹⁴⁷ or startups, where

¹⁴⁶ It's jumping ahead a bit but This is a key reason why it's important to a replicable institutional model.

¹⁴⁷ "I'm busy writing my thesis. Here's my lab computer - it has everything on it. Good luck!"

employees are notoriously mercenary and it can be surprising if someone stays more than three years.

However, it's also important to consider constraints that the organization will face early in its life that will hopefully relax over time. While a DARPA-riff is structurally set up to tackle long-term projects, it's probably not a good idea for the first projects to take 6+ years to show results. The reality of the situation is that a new organization will have no track record and so people will (rationally) have less patience with it. Organizational reputation matters. If funders, potential employees, and collaborators start to see the organization as unable to get things done¹⁴⁸ it can become a self-fulfilling prophecy.

9. Corporate labs enabled work in Pasteur's Quadrant.

There's not much to say here. This is the tightrope that a DARPA-riff should try to walk: things that are too researchy for a startup too engineering-heavy for academia.

A DARPA-riff will need to invert the normal relationship between an innovation org and its money factory

Research orgs are typically *subsidiary* to their their money factories. Bell Labs was subsidiary to AT&T, PARC to Xerox, DARPA to the DoD and Congress, academic labs to grant givers, etc. This relationship generates so many of the incentives that run counter to producing high-quality and eventually-impactful research. A lot of the specific friction can be traced back to mismatched Buxton Indexes—the length of the period over which an entity makes its plans.¹⁴⁹ Research just has a

¹⁴⁸ People absolutely run out of patience even if you set expectations up front. Based on popular sentiment, if Blue Origin had any funders besides Jeff Bezos, they would have lost funding despite the fact that they literally have tortoises on their coat of arms.

¹⁴⁹ The Buxton Index comes from E.W.Dijkstra in *The Strengths of the Academic Enterprise*

longer time signature than the natural beats of elected officials, quarterly earnings, or LP fundraising rounds.

What if you invert this relationship? Instead of the revenue generating entity (money factory) being the ‘parent’ org, instead the innovation org becomes the parent.

First, it’s worth asking what the distinction between a parent and a child or subsidiary organization means. Ultimately it’s about power — whose priorities take precedence? In the most extreme sense, if there is a choice between an action that will destroy one organization or another, which gets priority? It’s uncomfortable to talk about inter-organizational power in the same way that it’s uncomfortable to talk about what will actually get you fired or promoted in an organization.¹⁵⁰ This power usually takes the form of contracts if the children are external to the parent and budgets if they are internal. To some extent the power is simply a function of which organization came first. If a parent organization could survive before its child came along, it’s a clear proof point that it can survive without the child.¹⁵¹

It’s tempting to say “ah yes! Inverting the relationship between innovation orgs and money factories will solve the problems!” but that is magical thinking. Inverting the relationship won’t solve the mismatched Buxton Indices, it will just change which side exerts power. That relationship will create a *different* set of constraints on what happens and what can exist. The inversion is rare enough that there aren’t many case studies but one could imagine projects never shipping because of perfectionist thinking or a hesitance to kill projects that aren’t going anywhere.

It feels slightly nonsensical to not have the main profit-generating entity as the parent organization because otherwise, where does operating capital come from before the money factories are created? Let’s call this the ‘bootstrapping problem.’ The challenge becomes finding a *transient* source of money that

¹⁵⁰ Peter van Hardenberg poetically names this an organization’s “Reaper function.”

¹⁵¹ Though of course circumstances could have changed to make this no longer true.

simultaneously puts few enough constraints on the organization that it can do the work to create cash-generating children while at the same time not warping its incentives towards generating cash-children as soon as possible at all costs.

There are examples of organizations that have pulled off the inversion with varying degrees of success. Organizations like Flagship Pioneering, Ink & Switch, Idealab, and TandemLaunch are all both the parent organization and the innovation organization while the companies that they spin out are the money factories. If you squint, Howard Hughes Medical Institute also fits this pattern - its funding comes from its foundation, which is the subsidiary organization. HHMI is especially noteworthy because unlike the others, its survival isn't contingent on spinning out companies. As we've noted, spinning out companies is a double edged sword. On the one hand companies are a powerful technology diffusion mechanism but on the other hand, there is a significant class of innovations that would create drastically less value for the world if their value had been captured by their creators.

How did each of these organizations address the bootstrapping problem? Flagship Pioneering, Idealab, TandemLaunch and their ilk — ‘researchy startup studios’ — are explicitly out to create high-growth, venture-backable startups. This goal allows them to bootstrap with for-profit investment capital. It also incentivizes them towards projects that can become high growth startups on a short timescale. Ink & Switch has bootstrapped with a combination of sources — investment, ‘friendly investment’,¹⁵² and consulting with companies on the implementation of different areas of their research. Howard Hughes Medical Institute cheated by starting already sitting on top of a massive endowment. The

¹⁵² ‘Friendly investment’ is what I would call investment from non-professional investors that is some superposition of philanthropy and investment. That is, unlike professional investors who implicitly or explicitly want to see a certain level of growth on a clear timeline, friendly investment is more like an option - if the asset becomes worth something, awesome but if not disappointing but 🤷. Of course, friendly investment always runs the risk of becoming unfriendly at any time - the classic example is entrepreneurs who get their first startup money from friends and family who have no experience with startup investing. A lot of angel investing could be considered ‘friendly investing.’

upshot is that there isn't a well-worn path for an innovation organization to get to a point where its operating capital can consistently come from subsidiary organizations if it wants to work on things that might not want to become high-growth startups. Charting different possible approaches to the bootstrapping problem is the focus of the next section.

How does money work in an DARPA-riff?

Research is expensive. DARPA's 2020 budget was \$3.6B.¹⁵³ In the long run a DARPA-riff that works on atom-based technologies is going to need a lot of money. How money works is a major force that shapes incentives and organizational capability so it is worth taking very seriously.

When considering the pros and cons of funding sources it's important to think about them in terms of how much they could potentially blow the organization off of its 'ideal' course of action. This isn't to say "money taints everything!" It is to say that money comes from people external to the organization who have their own agenda (otherwise they would just be part of the organization!¹⁵⁴)

This section explores both the many-forked possibility path for making money as well as meta-considerations surrounding it.

Could DARPA have a positive return?

Spoiler alert - I suspect the answer is no, given any currently existing value capture mechanism.

Profit is important for both practical and ideological reasons. One of the most important reasons is that profit enables institutions to be self-sustaining. In steady state, could an organization riffing on DARPA sustain itself or would it constantly need external money?

¹⁵³ See Department of Defense Fiscal Year (FY) 2020 Budget Estimates.

¹⁵⁴ Of course, individuals within organizations also have their own agendas.

It's not an unreasonable question — technology eventually needs to be commercialized and diffused in order to actually impact people's lives, which means that eventually it will end up as products on the market. However, it's entirely possible for the product to be so many steps removed from the people who created the original technology that you can barely see the connection, let along propagate value back to the technology's creators.

One way to get a handle on the question of whether a DARPA-riff could be self-sustaining would be to ask "would (actual) DARPA have had a positive return?"

Answering this question is tricky. DARPA doesn't do any research in house, so all research they fund was ultimately done by another organization. Externalized research has a sparse paper trail and doesn't have an obvious value capture mechanism. Even if DARPA was able to directly capture value from its programs, those programs may have the impact they did *because* DARPA didn't try to capture their value. On top of all of that, valuable things that can be traced back to DARPA-funded work are usually combined with a lot of other work. This research admixture means that even if you could track down ever dollar of sales associated with DARPA research and assume that DARPA would have been able to capture some of it, a fixed percentage of that amount is a terrible assumption.

Given these hurdles, a more tractable question is "how much monetary value has DARPA spent and created?" Even answering this question is hard — DARPA is rarely the first or last organization to touch a project and on top of that it's unclear how much their help contributed. Given an accounting of created value, we could then play around with "given different scenarios, how much of that created value could have been 'captured' and what would have been the effects?"

The numbers I'm going to throw out are sketchy back-of-the envelope calculations — in reality, a rigorous accounting of DARPA's full-scale output is a full research project. I'm surprised the government hasn't commissioned such a project, but I've searched high and low to no avail.

DARPA's cost is straightforward. DARPA's budget in 2020 was \$(2020)3.6B. Its budget has been roughly stable since its founding except for a massive initial spike because it was originally responsible for what would eventually become NASA and various missile programs. Assuming a constant budget, DARPA has cost \$(2021)221B over its 63-year history.

DARPA's output is complicated. The simple place to start is with the stock market: which companies on the S&P 500 owe their existence to DARPA technology and how much are they worth? You can see the list and associated calculations [here](#). The total comes out to approximately \$(2021)11,200B¹⁵⁵ — two orders of magnitude bigger than DARPA's entire cost. Obviously, DARPA isn't responsible for anywhere near all of that value, but also unclear what percentage to use. YCombinator's (sketchily-estimated) 1.7% ownership at Doordash's IPO¹⁵⁶ is a hand-wavy anchor point. Despite ultimate responsibility for a lot of core technology, let's be conservative and say DARPA is 'owns' an order of magnitude less at 0.2%. This would come out to a DARPA ownership of \$22.5B — an order of magnitude *less* than the total amount that was spent on DARPA. This is surprising!

Of course, this calculation is a massive underestimate: it's only counting public companies, it's ignoring any dividends that they may have paid out, is giving DARPA a tiny ownership, and I'm being very sloppy about what's attributable to DARPA. At the same time, I'm skeptical that private companies (there are many highly-valued autonomous car companies especially) or dividends would make much of a difference or that DARPA would own much more, especially given existing value capture mechanisms. I don't have strong evidence, but I strongly suspect this \$22.5B number is anywhere representative of the amount of value DARPA has created. The reality may truly be that most of

¹⁵⁵ I only took the top 20 companies because company valuation roughly follows a power law so the bottom 200 will probably add up to be much less than the top 20.

¹⁵⁶ See Running The Numbers on Y Combinator's Best Year Yet.

the value that DARPA has created isn't easily measurable. How much modern activity, joy, and wonder is enabled by GPS and the internet? What has been built on top of obscure research funded by DARPA? How many careers has it touched? The trick is that if you can't measure value, you certainly can't capture it.

Is there a minimum effective budget to generate outlier results

It is important for big things to start small — many attempts to replicate DARPA's results start too big¹⁵⁷. So it makes sense to ask: is there a minimum effective budget to generate outlier results? Can you start with a tenth of DARPA's budget? A hundredth? A thousandth?

There are two lower budget bounds that you'll run into: the number of programs you need to run to get a single 'successful' program and the budget per program for it to have a viable shot at that success.

A back of the envelope calculation suggests that if each program generously has a 10% chance of success (DARPA has a 5-10% program success rate)¹⁵⁸ then you need to run at least seven programs to give yourself a 50% chance of at least one program being successful¹⁵⁹. Whether these programs are run in series or parallel depends on available funding. However, if it takes five years to run a full program, running them in series could take a decade or more (assuming several unsuccessful programs are terminated early) to even demonstrate organizational feasibility. And this isn't "ten years to a \$1B IPO" — it's "ten years to getting a really promising technology out into the world that given another decade might produce a measurable impact." Given the absurd timeline of running programs

¹⁵⁷ See the section Most DARPA clones start too big or with heavy process from Why Does DARPA Work?

¹⁵⁸ See DARPA—Enabling Technical Innovation

¹⁵⁹ See numbers appendix.

in series, running several programs in parallel is important. The necessity of both starting small and running several programs in parallel to even demonstrate feasibility is one of the core tensions inherent in riffing on DARPA.

The question of minimum program budgets is tricky. Let's first look at some comparison numbers. In the years 2018-2020 among DARPA programs not focused on assembly and production, the minimum budget was \$2m, the maximum budget was \$31.4m and the average budget was \$12m.¹⁶⁰ The ARPA IPTO directorate that midwifed the personal computer started with a budget of \$ (2020)47M in 1962.¹⁶¹ ARPA-E vacillates between ~\$200-\$300M/year, has about 50 programs running at any time, which comes out to roughly \$4m to \$6m per program. IARPA's budget is classified. Could a program go below a couple million dollars a year and still be effective? There are arguments both ways. On the one hand you could argue that surely bureaucratic inefficiency is pushing those budgets higher than they need to be. On the other hand, it might be that the sort of work that it's most valuable for a DARPA-riff to do needs a relatively large budget, otherwise it happen in other institutions. Without strong reasons to believe otherwise (which we don't have), it seems irresponsible to assume that you could run impactful programs for significantly less than DARPA. The saving grace is that programs don't need to start at full-budget but can ramp up over a year or two.

So is there a minimum effective budget to generate outlier results? Absolutely. The number depends heavily on how often you expect programs to fail and the type of work those programs will entail. There are different strategies you could use to both accept the reality of a minimum effective budget and at the same time start as small as possible to avoid money poisoning. In part III, I'll argue that (one) feasible approach is a combination of explicit evolutionary forms and tranches.

¹⁶⁰ See Department of Defense Fiscal Year (FY) 2020 Budget Estimates and in spreadsheet form [here](#)

¹⁶¹ See Arpa Does Windows: The Defense Underpinning of the PC Revolution

Income

DARPA-riffs need free cash flow. You need money to pay people, rent lab space, purchase equipment and reagents, hire lawyers keep regulators from shutting you down, and possibly pay to rebuild anything that you blow up in the process. Figuring out where this money comes from is a narrow path between the Scylla of external constraints and the Charybdis of organizational death. Hew too far towards any established funding source and the constraints it imposes won't let you do anything different from a normal startup, nonprofit, consulting firm, etc. Hew too far the other way and all that freedom will be worthless because you'll need to shut down the organization. The rocks of history are littered with the smashed wrecks of organizations that failed to run this gauntlet — Magic Leap is a prime example of the former and Dynamicland of the latter. There are many others but I avoid being mean on the internet so I'll leave you with a pregnant silence.

Three big considerations for any income stream are whether it is a transient source that could be used by the organization in its early days or could provide steady state income, what order of magnitude you could expect it to provide (thousands? Tens of thousands?), and how replicable it is. The last point is worth considering because ideally DARPA-riffs would become a replicable institutional model.

This section will walk through potential directions at the many-forking juncture that asks “where does free cash flow for an ARPA-riff come from?” Or in other words, what are the different tools a DARPA-riff can use to build its money factory?

Technology Licensing

Licensing fees from patents are a straightforward way to make money for an organization creating technology. Licensing patents lets you focus on creating the technology rather than commercializing it. Licensing also enables you to create

an organization that can generate income from outputs that wouldn't necessarily make good startups — processes or technologies that benefit from the existing manufacturing, expertise, and supply chains of large organizations.

A few organizations make a significant chunk of money from licensing revenue. Between 2008 and 2012 IBM made over \$1B / year in licensing revenue.¹⁶² Intellectual Ventures' primary business model revolves around licensing. Universities are another example. [From an article on licensing revenues in 2015:](#)

Northwestern University topped the income list, at \$360 million, a number that was boosted by the sale of its remaining royalty rights to a compound used to control neuropathic pain that is marketed as Lyrica. Other big earners in the survey included New York University, \$215 million; Princeton, \$142 million; Columbia, \$115 million; the UC system and Stanford, each about \$108 million; and the University of Texas system, \$49 million.

All of these examples do some good research work! However, there are some flies in the ointment. The majority of university licensing income is from drug licensing deals. Both IBM and Universities spend far more on research than they make in licensing revenue. Most technology transfer offices are not profitable.¹⁶³ All the universities in Canada combined only do \$62M in licensing revenue.¹⁶⁴ And the steps Intellectual Ventures needs to take in order to get paid have earned them a rather unpleasant reputation.

At the end of the day, patents need to be enforced through the law. The only way to do *that* is through lawsuits. Ideally people will proactively pay licensing fees but often they do not, either through maliciousness and negligence. As a result,

¹⁶² See If Patents Are So Valuable Why Does IBM's Intellectual Property Revenue Continue To Decline

¹⁶³ See Benchmarking of Technology Transfer Offices and What It Means for Developing Countries

¹⁶⁴ See Canada needs a national overhaul of university IP policies

any organization depending on patent licensing fees needs to also have a staff of (expensive) lawyers constantly monitoring for patent violations and ready to sue to enforce patent rights. This unfortunate necessity is why Intellectual Ventures is maligned as a patent troll. It might be possible to decrease violations by using non-exclusive licenses. However, non-exclusive licenses are usually worth much less than exclusive licenses, even counting the fees from multiple licensees.

Heavy dependence on licensing fees tends to warp incentives. IBM's internal status game is on publishing patents — people are promoted based on the number of patents they file rather than getting anything to work. Patents as an accomplishment metric leads to projects being abandoned once they get to a patentable state. I've experienced this personally at other organizations as well.

Licensing obsession also leads to a conflict of interest between researchers and the organization. If a researcher thinks they're on to something valuable, they're incentivized to keep it secret, leave the organization, and set up their own venture to capitalize on the thing they've created without paying licensing fees. Such a surreptitious course of action is totally rational — licensing fees are a deadweight loss on any organization that has to pay them, especially a startup. This effect seems especially likely to happen if some or all of the research is externalized.

On top of the organizational problems that a licensing-fee-based income strategy can introduce, patents add a lot of friction to innovation. There are examples throughout history of patents keeping systems that depend on multiple patents held by different entities from being built. Aviation is arguably one of the most notorious examples. Patents hamstring the entire class of innovations that would create drastically less value for the world if their value had been captured by their creators.

This isn't to say that a private DARPA should *never* patent anything. However, it would need be on a case-by-case basis and have a strong reason why the patent would be in service to the technology eg. enabling it to raise further financing where otherwise it could not. The upshot is that while patents could be a way to

spuriously create income, licensing is a bad primary income strategy for a private DARPA.

Spin-off Companies

Another obvious way for an organization creating technology to make money is by spinning off startups to commercialize that technology, investing in them, and using the revenues from liquidity events like acquisitions or IPOs to fund operations.

One might think the prevalence of startup studios is a strong proof point in favor of the spin-off. However, most startup studios (rationally) focus on company ideas with low tech risk — well-trodden software paths, direct to consumer models, etc. There are, of course, notable exceptions — [Idealab](#), [Flagship Pioneering](#), Deep Science Ventures, and TandemLaunch¹⁶⁵ to name a few. Each of these organizations does take on significant technical risk but the need to create and capture venture-scale returns limits the sorts of work they can do. In the vast majority of cases, research-y startup studios are constrained to work on classes of problems that have clear cut sales channels and product pipelines like pharmaceuticals (and other medical technology), software, or electronics. They also primarily focus on work that has been de-risked by a single academic lab.

An income stream based on spinning off companies has a number of advantages. As anybody who has even heard the words “Silicon Valley” knows all too well, equity in high growth startups can have truly massive upside. The money from a single successful spin-off can fund many others. Spin-offs could also help align incentives. If a program manager could potentially become a CEO of a company, a DARPA-riff could possibly avoid situations where smart, talented people say “that sounds good, but I could make much more going off and starting a startup.” Another advantage of spin-offs is that in the long run, research affects

¹⁶⁵ For more on Deep Science Ventures and Tandemlauch see Episodes 19 and 30 of the Idea Machines podcast respectively.

the world in one of two ways: knowledge or products / processes. Companies are entities whose incentive is (to a large part) to diffuse products or processes as far and wide as possible. So if you're creating new technology that you want to get out into the world (as opposed to sitting on a shelf) creating a company around it is probably a good idea ... eventually.

At the same time, spin-off-based revenue introduces a number of constraints. The biggest is the danger that if every program is aimed at becoming a company, the organization will be incentivized away from things that would create drastically less value for the world if you try to capture their value. Focusing on spin-offs could potentially warp incentives to work on things that could produce short-term wins and things that would be able to raise money easily once they've spun out; spin-outs will probably only succeed if they can raise follow-on capital and only turn into income when they liquidate so fundable, shorter-term wins are a more predictable income stream.

Spin-offs as a primary income source also creates a liquidity problem. Even if you avoid all the incentive problems and manage to do maximally important work that could also become a valuable companies, given the state of private and public markets, those companies probably won't generate a return for many years. In that situation, you need to get operating capital from outside sources — presumably equity investors. These investors will want their own above-market returns, which will apply pressure that can propagate all the way back and warp which programs you undertake in the first place. Pressure from the same sorts of investors who invest in VC funds will create the same pressure that those funds are under. Even with understanding, long-term investors, there would need to be a justification of how any program could become a company. But there are some pieces of a future that maximizes human capability and wonder that might just not make good companies!

Case Study: Intellectual Ventures

[Intellectual Ventures](#)(IV) has a unique structure that combines income from technology licensing and spin-offs.

Three of IVs four divisions (Invention Science Fund, Intellectual Ventures Lab, and Deep Science fund) are focused on creating internal patents and spinoffs. So far the companies they've created have raised \$700M in venture funding¹⁶⁶, which in roughly puts their total value between \$5-10B. Assuming IV did not take an equity share too large to dissuade investors (<20% perhaps?) that would put their share at \$1-2B - far less than the \$5B invested externally and who know how much originally put in by its founder, Nathan Myhrvold. It's likely that it's heavily subsidized by the patent buying and licensing business (ie. The troll part.)

Could IV pay for itself without the patent troll part of the business?

Traditional Philanthropic Funding (Donations)

Donations without any hope of monetary return seems like an obvious way to fund things that need to be public goods in order to have an impact.

The biggest advantage of donation-based income is that an organization can unabashedly work on things that won't generate a profit. The ability to decouple from profit is a fairly self-explanatory boost in freedom of action. A more subtle benefit is that external researchers and organizations might be more willing to collaborate because (non-political) nonprofits are perceived as having fewer ulterior motives.

There are several successful or promising research organizations funded entirely by traditional philanthropy. Among them are the [Allen AI Institute](#), the [Howard Hughes Medical Institute \(HHMI\)](#), the [Santa Fe Institute](#), the [Gates Foundation](#), the [Chan Zuckerberg Initiative](#), and [Wellcome Leap](#).¹⁶⁷ The latter three blur the line between 'research funding' and 'research execution' but so does DARPA. While not a research institute per se, the [X-Prize Foundation](#) is a good example of

¹⁶⁶ See <https://www.intellectualventures.com/spinouts>

¹⁶⁷ It's worth noting that of these, only the Santa Fe institute does not have a large endowment.

a philanthropically funded innovation organization that pioneered a weird new model. At the same time the fate of Dynamicland¹⁶⁸ illustrates that doing paradigm-shifting, well-regarded, visionary work is not sufficient to bring in enough donations to fuel an organization.

While not being beholden to investors or customers creates a lot of freedom, donation-based income comes with a whole raft of constraints.

People don't like giving away money. Furthermore, the reticence tends to be supra-linear with the amount of money (regardless of the total wealth fraction that money represents). Fundraising for *anything* is hard; even more so when there is no potential return on that money.

While, companies (hopefully) get to profitability after a few fundraising rounds, donation-based income never escapes the fundraising game. Investors like VC or hedge funds are in a similar situation but the differences between them and charities is illustrative. The key difference between a donation- and a investment-based fund is that the latter has a clear metric (usually internal rate of return (IRR)) that it can use to encourage future investment, despite the whole "past performance is not indicative of future results" rigamarole. A donation-funded organization that tries to have an equally clear metric will inevitably fall victim to Goodhart's Law¹⁶⁹. Raising money without a metric means that you're completely at the whim of people's feelings. Bluntly, most people give money to philanthropy to make themselves look and feel good.

Raising money without a metric forces an organization do a lot of 'donor management.' In addition to time and money (charity balls anyone?) you need to show progress every time you ask for money — once or twice a year.¹⁷⁰ Because most donors can't or don't want to get into the weeds, these progress reports

¹⁶⁸ While not official, my understanding is that after struggling to find funding Dynamicland closed its doors in 2020 (in part due to COVID-19 but that was only final nail.)

¹⁶⁹ "When a measure becomes a target, it ceases to be a good measure"

¹⁷⁰ Charities can't lock in commitments the way for-profit funds can

need to take the form of shiny demos which can end up distracting from the core mission or worse, biasing the entire organization towards programs that create shiny demos.

Donation-based income is especially vulnerable to economic downturns. Economy-coupled funding is especially dangerous for long-term research because most researchers are not fungible. You can't just lay off researchers and hire them back when there's more money. Labs will find different sources of funding and talented people will always be in demand somewhere. Even if you could, programs would take a massive hit — the nature of research makes it impossible to start back up exactly where you left off.

It's also important to consider the scale of donors and donations and the associated tradeoffs. On one end of the spectrum are individuals giving a few dollars and on the other hand you have an organization like Wellcome Leap or the Allen AI Institute funded by a single massive foundation.

A brief aside on terminology — So far we've managed to avoid the terms "nonprofit" and "charity" when talking about donation-funded organizations. These terms are tricky because they have both colloquial meanings but are also legal designations that come with a slew of implications — especially around taxes and allowed activities. These implications warp behavior enough that it would be sloppy to flip between using (little-n) nonprofit — an organization that is not seeking to maximize profits — and (big-N) Nonprofit — an organization incorporated under section 501(c)(3) of the US tax code. To avoid confusion for both you and myself, I'm going to try to use both "Charity", "Nonprofit", and "Foundation" in the legal sense and always capitalize them to occasionally remind you of that usage.

The many-small-donors end of the spectrum tends to push organizations towards either short-termism or constrained sizes. You'll notice that Charities that depend on many small donations are constantly asking for donations — everybody has seen Wikipedia's yearly banner; and many of us remember asking for pennies for UNICEF on Halloween. Even if they aren't cash-strapped,

Charities need to get 33% of their income from ‘small donors’¹⁷¹ in order to maintain their legal status (and associated tax benefits for themselves and donors)¹⁷². “Yeah, we’re still struggling with the same experiments we were last year” is a common occurrence in research but isn’t a great way to get people excited about donating, nudging charities towards shorter time horizons and flashier results.

Small donors also make funding even more unpredictable than the economy-coupling you would expect from any donation-based funding. If you have a few large donors you can stay in touch and at least get a heads up on whether they’re thinking about donating this year or not.

There are a few ‘independent researchers’ whose incomes depend on donations through platforms like [Patreon](#) or [Github Sponsors](#). These folks forgo the tax-benefits and funding requirements of being a legal Charity in favor of longer-term funding stability. Empirically, this approach seems to be able support an individual but not a team or work that has large capital requirements. However, this corner of the funding maze seems under explored and might be able to be productively combined with eg. membership models.

The ‘few, massive donors’ end of the philanthropic spectrum is dominated by large Foundations. Outside of a few Foundations that are directly controlled by their founder, most are controlled by professional managers. Professionally managed Foundations usually fall into the asymmetric career risk trap — managers are punished for funding something risky that fails but get relatively little upside from funding something risky that succeeds. As a result, Foundations want to see clear metrics, updates, and derisking on the money they give out. In reality it’s shockingly hard to get professionally managed foundations to do weird new things.

Given the issues with small donors and professionally managed Foundations, a common refrain is often: “Just find a few wise, long-term thinking billionaires

¹⁷¹ ‘Small’ here is <2% of total funding

¹⁷² Howstuffworks has a reasonable explainer

and you're set!" This is lazy thinking for several reasons.

1. There are only 2095 billionaires in the world as of 2020. Being a billionaire doesn't shift your chances of being wise and long-term thinking too much 😱 so there just aren't that many wise, long-term thinking billionaires.
2. Wise, long-term thinking billionaires already have a bajillion people banging down their door for money because everybody else has the same thought.
3. Partially as a result, many wise, long-term thinking billionaires have already bucketed their philanthropic dollars into a foundation.
4. Being worth a billion dollars on paper doesn't mean that someone has a billion dollars in a bank account they can actually give away.
5. Perception of money isn't strictly a fraction of how much money you have — tens of millions of dollars is a lot of money, period.
6. If a single person is bankrolling an entire project they (for good reason!) tend to feel some ownership over it. Depending on the person, this attitude may be problematic for an organization that benefits from opacity and independence.¹⁷³
7. A single source of funding means a single point of failure for the organization if the funder decides to pull out — a not-uncommon occurrence. The upshot is not that it's a bad idea to consider funding from wealthy individuals but that it is a bad idea for that to be the only plan.

(Legally designated) Nonprofits are subject to a raft of requirements and constraints about what they can and cannot do without losing their Nonprofit status (and thus loading donors with a host of unexpected taxes). The yearly small-donor requirement on Charities is just one of them. The existence of multiple, ill-defined actions that a Nonprofit is not allowed to take is especially

¹⁷³ See the section Opacity is important to DARPA's outlier success from Why Does DARPA Work?

problematic for an organization where it's unclear a-priori what it's going to take to achieve an R&D goal and wants to be able to move quickly.

People who donate money generally expect that money will not be used to enrich any individual. However, it's nebulous a priori which research should produce public goods vs private goods. Sometimes the best way for a technology to have an impact is as part of a private, for-profit enterprise. This tension would need to be treated carefully.

Philanthropic Investing

Philanthropic investing sits in a grey zone between a donation and an investment and is definitely worth investigating for a DARPA-riff. 'Philanthropic investing' can happen in one of two ways: when a Foundation makes a return-producing investment out of the money that it's spending on its 'charitable mission'¹⁷⁴ or they make a risky investment out of their endowment¹⁷⁵. In both cases, the foundation is bending the normal rules in order to make an investment that's aligned with the Foundation's mission. Under normal circumstances, Foundations can only use their 'programmatic' money to give grants to 501(c)(3) organizations ('big N' Nonprofits¹⁷⁶) and endowments can only make 'non-risky'¹⁷⁷ investments.

The potential to tap into Donor-Advised Funds is another reason for a DARPA-riff to think about philanthropic investing. Donor-Advised Funds (DAFs) are like Foundations' weird younger siblings: both organizations are vehicles for wealthy folks to philanthropically spend their money; foundations

¹⁷⁴ Foundations are legally required to disburse at least 5% of their assets towards their charitable mission every year.

¹⁷⁵ Legally, foundations aren't allowed to make 'excessively risky investments.'

¹⁷⁶ Are you starting to see why all these legal shenanigans matter?

¹⁷⁷ Yes, I'm using an absurd number of quotation marks - this is yet another artifact of the fact that There are no first principles for legal structures! All of these words are legal designations, but it's nebulous which things fall in and out of them.

give their founders more control but have more overhead and restrictions; DAFs require less initial capital so they appeal to many wealthy-but-not-name-you-know-wealthy folks. For our purposes, one of the biggest differences is that unlike Foundations, DAFs don't have a payout requirement. As a result, there is a lot of earmarked-for-charity-eventually money sitting around collecting dividends.¹⁷⁸ The hypothesis is that (in part because they have smaller staffs to evaluate non-money metrics) philanthropic investments are a way to connect DAF money to research projects.¹⁷⁹

A DARPA-riff's ability to take advantage of philanthropic investing is a complex dance with its own structure. Return-producing investments made from a Foundation's charitable spending (also called program-related investments or PRIs) can happen through several mechanisms. These mechanisms include 'recoverable grants' directly to a for-profit organization or funding through an intermediary Nonprofit.¹⁸⁰ If a DARPA-riff was organized as a for-profit organization it could get income directly from a recoverable grant; if it was organized as a Nonprofit, it could only act as an intermediary for research that had already budded off into its own organization. Depending on the exact legal requirements about these different funding vehicles, direct grants or loans may require programs to be separate entities from the umbrella organizations and thus may be appropriate for more advanced (post-seedling) programs.

It's likely that philanthropic investing will primarily need to be done on a program-by-program basis. The funding needs to align with each funder's specific mission so there will also most likely be a lot of fiddling involved to match programs with funders that will require constant inputs of time and effort. Many programs might not go to align with *any* causes because charitable

¹⁷⁸ See *Zombie philanthropy: The rich have stashed billions in donor-advised charities — but it's not reaching those in need (paywalled)*.

¹⁷⁹ See *Donor-Advised Funds: an underutilized philanthropic vehicle to support innovation in science and engineering* for many more details

¹⁸⁰ For a more in-depth flowchart of the different ways PRIs can happen based on different constraints see *Foundations: exploring the emerging practice of philanthropic investing to support innovation in science and engineering*

causes tend to be focused on ‘problems’ and DARPA-like program tend to focus on ‘more awesome’ rather than ‘less bad’. Program-by-program funding would require the umbrella organization to find a different source of funding for things that aren’t full fledged programs like seedling efforts that could possibly go nowhere. Earmarked philanthropic investments would also limit the fungibility of funds between different programs that allow DARPA to take on more risk.¹⁸¹

This is all rather speculative. In contrast to the fairly-well-trodden paths of technology licensing, spin-offs, and donations, philanthropic investing is a relatively new thing. The core hypothesis is that DAFs and Foundations would be more likely to give if there was some chance that they could recover some of the investment (so that they could fund more good things.) [Prime Coalition](#) is one of the pioneers in the area. They use recoverable grants to do equity-funding of for-profit companies. If you squint at it, a private DARPA could have the same relationship to Prime Coalition’s structure as a startup studio does to a VC firm.

Membership Model

A private DARPA could generate some free cash flow from a membership model where other organizations pay to be part of a ‘business network’ where they get access to special events, first crack at research, and help implementing it. The [MIT Media Lab’s funding consortium](#) is the most prominent (and successful) example of this approach.

Relative operational freedom (if managed correctly) is potentially the biggest upside of the membership model. It can provide short-term liquidity without needing to provide either a monetary ROI like an investment while having perhaps less volatility than charity if you can convince organizations to buy

¹⁸¹ See the section [It is relatively easy for DARPA PMs to re-deploy funding from Why Does DARPA Work?](#)

longer-term memberships.

In addition to money, the membership model could help more technology get out into the world. One of the most failure-prone points in technological development is when a technology needs to move from an organization that focuses on R&D to one that focuses on marketing and manufacturing. Having large organizations with marketing and manufacturing capability in the loop and already ‘bought in’¹⁸² could reduce a lot of friction. A membership model could help shepherd frontier technologies with bespoke sales channels in the same way that some pharma investors’ main role is as a smooth pipeline through the FDA approval process to acquisition. Short of a program literally *becoming* the marketing and manufacturing organization for the technology it worked on¹⁸³, a DARPA-riff needs to interface with external orgs to smooth transitions to maximize its impact, with or without a membership model.

A less-tried version of the membership model would extend it beyond big companies to individuals as well. This extended membership model would almost be a hybrid between charity and providing a “product” in the sense of enabling people to be a part of something bigger. How many people would sign up to be a SpaceX member? Probably a lot! The closest analog is probably independent researchers who fund their work through Patreon.¹⁸⁴ There are many unexplored and potentially exciting paths surrounding individual memberships — point systems, tying donations to individual units or work or purchases, community participation in the research itself. The biggest open question is whether an agglomeration of small individual memberships can fund more than a small team. For all the people who would be excited to be SpaceX members, would they even fund a single starship test launch?

Unfortunately, the membership model will probably be unable to fund

¹⁸² Of course, members of a consortium are not automatically bought in to everything that comes out of it.

¹⁸³ Spoiler alert: Speculatively, this is a good idea!

¹⁸⁴ See *Reflections on 2020 as an Independent Researcher* for a more in-depth description of this approach.

operations on its own. The MIT Media Lab is probably the most successful organization using this model and its operating budget is only \$45M¹⁸⁵ (approximately four average-sized DARPA programs). It's unlikely that a private DARPA could do much better. Additionally, expenses with fuzzy ROIs are often the first to be cut from organizational budgets during lean times so membership-based income won't be particularly stable. Notoriously, The Media Lab's incentives drive it to constantly put out shiny demos that ultimately go nowhere. Member management — putting on events, making half-finished projects sounds exciting — could take up time and resources.

These constraints mean that a membership model is certainly a promising income stream but should probably be coupled with other income sources.

Government Grants and Contracts

While 21st century riffs on DARPA should be private, the government gives both grants and contracts to plenty of private organizations. Would a private DARPA be able to take advantage of them?

Government Grants

The advantage of government grants is that they come with relatively^{*186} few strings attached. Governments don't care about either direct monetary ROI and can nominally act on long timescales. The government can actually capture value from public goods, so their incentives are nominally aligned¹⁸⁷ with an

¹⁸⁵ See <http://catalog.mit.edu/mit/research/mit-media-lab/>

¹⁸⁶ Beyond reporting requirements which as we'll see are thick strings indeed.

¹⁸⁷ Of course, the government is composed of individuals who all have their own agendas and geographies.

organization whose goal is to create impactful technology. Governments also deploy a massive amount of money so the operating budget of a DARPA-riff isn't outrageous compared to other expenditures.

While multi-year government grants are more stable than donations that you need to worry about on a yearly or semi-yearly cycle, they are often one-off events. As a result, an organization that depends on government grants for income needs to spend a significant chunk of time applying for and administering those grants.

There's also a question of specifically which grants we're talking about. [Small Business Innovation Research \(SBIR\) grants](#) are one obvious answer. These grants are explicitly designed to fund private R&D that has the potential for commercialization. However, the money from SBIRs is often barely worth the time spent to get it — “Although the applications [for less than \$1M] may only be 10 or so pages long, companies should budget at least 80 full-time hours to complete them.”¹⁸⁸ So SBIRs are ... less than optimal, meaning that it will be a job to even hunt down the right grants to apply for.

Despite coming with relatively few incentive-warping strings, government grants do come with plenty of restrictions on how the money can be spent. Most government grant money is earmarked — either to be spent on a specific project or for a specific purpose like salaries or equipment¹⁸⁹. Spending restrictions are clearly problematic when you want to be able to do whatever is necessary to get the work done. It's unlikely that a DARPA-riff would be able to get much grant money that wasn't tied to a specific program. You would also need to either contort what the organization is working on to align with existing government funding priorities or do a lot of work to convince people to shift those priorities.

Government grants also tend to either fall into the under \$1m or over \$100m range, both of which are suboptimal for a DARPA-riff. Going for non-standard grant

¹⁸⁸ See How to access ‘America’s Seed Fund,’ the \$3 billion SBIR program

¹⁸⁹ This is why unrestricted money is like gold for professors.

sizes will require a lot of work. You also need to convince people in the government that this weird institutional structure is important and worth funding. In other words, a lot of lobbying (or put more bluntly, rent-seeking) would be involved.

Government Contracts

Government contracts are like grants with more clearly defined deliverables. Unlike selling products, the payment is upfront and the ‘deliverable’ doesn’t necessarily need to be as polished and standalone as a product. The deliverable could be the output of a piece of research or a proof of concept.

[SRI International](#) is a great case study of an innovation organization whose income is primarily based on contracts. Arguably, SRI is co-responsible for many of DARPA’s triumphs — Douglas Engelbart did the work that resulted in the Mother of All Demos at SRI — and has spun off many valuable and well-regarded organizations.¹⁹⁰ At the same time, they’re also the prime example for the constraints below.

Contracts have advantages over both grants and equity funding. Research contracts can be much larger than grants while still enabling work that might not result in a product. There are fewer restrictions on how you can deploy money from contracts and at least part of that money comes in before doing the work so you can hire necessary people and buy equipment.

Depending on contracts as a primary source of income pushes an organization to increasingly resemble a government contractor despite lofty goals otherwise. Once you start playing the government contracting game, the path of least resistance is to *become* a government contractor. Even with this possibility in mind, it’s impossible work on government contracts without slight organizational warping. You need to spend time and effort paying attention to

¹⁹⁰ Including Intuitive Surgical, E-Trade, Symantec, etc. See this Wikipedia page for a full list.

which contracts are available, bidding for them, and administering them, so contracts aren't something that you can easily do occasionally and opportunistically.

Contracts are fairly unpredictable. SRI International, for example, has regular boom and bust cycles, where they hire many people to work on a particular project and then need to fire them when they no longer have contracts to work on. Unpredictability might be slightly less of a problem if more of the work is externalized, but it would still be problematic for long-term projects.

Obviously the biggest problem with contracts is that you don't have much control over what you're working on. In a best case scenario, you can find an existing call for something that you already need to do or a slight extension of that work. In a less-good scenario you either end up lobbying to get contracts for what you were going to do anyway or need to heavily shift what you're doing in order to win them. Again, the issue is that in order to maintain the organizational infrastructure to do the former, you'll probably need to at least some of the latter.

Despite all the pitfalls, organizations like [Otherlab](#) and many other small engineering firms have managed to get by on grants combined with contracts. However, their work often feels scattered and incomplete. The need to pursue different contracts and grants makes it hard to do the sort of deep long-term work that can lead to extensible breakthroughs.

Consulting

A DARPA-riff could sell program designs to large organizations and provide consulting services for implementing them. There are a number of organizations with a similar model; [Lux Research](#) charges tens of thousands of dollars for super high quality reports on technical areas and the ability to ask questions about them. If the hypotheses that program design will be a core part of a private DARPA, it will already be creating artifacts that larger organizations might be

interested in.

A consulting-based income stream has a number of upsides, especially early in a DARPA-riff's life. It could generate early cash before even implementing programs while maintaining relative operational freedom from investors or donors. If the program designs would be important and good for the world (as I would hope all them are!) it's a win-win whether or not the large organizations actually execute on the program. If they do execute on it well, it's a cheap demonstration that the program design process works. If they don't execute on it or execute on it poorly, you could still run the program later (assuming there is no reputational blowback to the organization or the program idea itself). Selling knowledge to large companies with manufacturing capability could potentially address frontier technology's bespoke sales channel problem. One could imagine a 'pipeline' where even if big organizations don't fully embrace program designs, they nevertheless start to acclimate to the ideas. In this way, there's a fuzzy boundary between consulting and the membership model.

Consulting-based income comes with a slew of downsides. The biggest downside is that consulting wouldn't provide enough cash to support a full-blown steady-state DARPA-riff. That being said, it could be a good way of bringing in income while demonstrating that program design can work. Consulting is a canonical trap for small organizations — clients can often pull you in unproductive directions and you need to spend a significant chunk of time on sales. Another danger is that this income stream would render a DARPA-riff particularly susceptible to working on 'hot' areas that are adjacent to high margin industries. This danger is particularly insidious because the most important things to work on may be unsexy areas adjacent to low-margin industries.

The Mitochondria Maneuver - Be an independent-ish arm of some existing money factory

As we've beaten to death by now, a private DARPA would fill a niche in the innovation ecosystem that was once filled by industrial labs. While new industrial labs created inside of corporations are probably not viable, it might be possible for an external organization to *become* a corporation's industrial lab while maintaining process and culture that supports high quality solutions R&D. This strategy towards long-term survival is "the mitochondria maneuver." Get absorbed into a larger organization while maintaining a permeable barrier and your own cultural and operational DNA.

The benefits are pretty clear, if they don't require neutering your organization to get them. Innovation organizations need a money factory and what better money factory than a big company that is already making massive profits? The budget from the larger org is effectively dividends from an endowment.

Arguably, [DeepMind](#) is a great example of the mitochondria maneuver.¹⁹¹ An AI research lab, they were acquired by Google in 2014 but unlike most acquired companies, they have maintained relative independence to do high-quality long-term work.¹⁹² As long as they continue to have hits every few years (AlphaGo, AlphaFold) and some of their work directly helps the mothership (some of their work makes Google's data centers more efficient) Alphabet seems willing to heavily subsidize their almost \$1B budget.¹⁹³ However, the fact that the Alphabet-DeepMind relationship meets the three criteria for healthy industrial labs should make you skeptical whether the mitochondria maneuver can be accomplished in other (lower-margin, less-clearly-aligned) domains.

¹⁹¹ Except for its majority-internalized research DeepMind is basically an AI-focused DARPA-riff.

¹⁹² DeepMind is technically an Alphabet subsidiary with its own leadership, budgets, etc.

¹⁹³ See DeepMind revenue dwarfed by \$649 million loss

While they're under the same constraints in the long run, a key difference between a mitochondrial organization and a home-grown lab may be the ability to skate where the puck is going. A 'normal' industrial lab needs to justify its existence from day one — 'stakeholder buy-in' and all that. By contrast, a mitochondrial organization only needs that justification *eventually*. DeepMind officially started in 2010 — two years before AlexNet acted like lightning to AI hype's thunder. It's hard to put yourself in the mindset of pre-2012 attitudes towards AI research, but the term 'AI winter' exists for a reason. Suffice to say, it seems unlikely that any corporation would have created a full-fledged AI-focused lab in 2010.¹⁹⁴ This framework suggests that a key requirement for the mitochondria maneuver to work is a strong thesis about how your DARPA-riff's focus will align with a high-margin corporation's priorities in the future.

A private DARPA cannot count on the mitochondria maneuver from day one. During its Odyssean wanderings, it will still need to depend on one of the other income sources. However, some income streams might be more conducive to eventually getting absorbed than others — a membership model or programmatic consulting could both become more and more serious until two organizations decide to go steady and consummate the relationship.

Attempting the mitochondria maneuver seems fraught with failure modes. It basically requires predicting the future, which is hard. (Though easier if you're doing that by inventing it.)¹⁹⁵ The steps an organization might need to take to survive might make it unpalatable for a corporation. The opposite could also happen! And while DeepMind has done well, history is littered with organizations that thought they were going to become their acquirer's R&D arm but the fates had other things in store for them.

¹⁹⁴ Google Brain's timeline corroborates this: it started as a small project within Google X in 2011 and only became a full organization around 2013.

¹⁹⁵ "The best way to predict the future is to invent it" — Alan Kay

An endowment that funds operations out of the interest on a large principle seems like the ideal funding mechanism for a private DARPA

Ultimately, a private DARPA needs to be autocatalytic. The only way for it to be able to plan and act over long timescales is to fund operations through predictable, long-term income that doesn't distract from the core goal. Unfortunately, these criteria rule out many funding sources for 'steady-state' operations. One of the only ways to satisfy them all may be to build an endowment.

Most of the income streams we've explored aren't suitable to fund a private DARPA over the long run. For income to be predictable, it can't disappear at the whims of either a few individuals or large mobs — so grants, philanthropy, and contracts aren't viable as steady-state funding sources. Technology licensing and membership models can't generate enough revenue to fully fund steady-state operations. Creating and selling products or services can create predictable long-term income. However, it seems hard to avoid sales-based-income distracting from the core goal when that goal is in part to enable things that are not product-focused enough for startups. Similarly, it would be very hard to avoid an investment-in-spin-offs-based income stream creating incentives to spin off more things with ever bigger, nearer-term exits. Spin-off-based income *might* work in steady-state, but to count on it alone would be to enter the same incentive regime as VCs. The mitochondria maneuver is one explicitly steady-state option, but it requires a fairly narrow set of conditions.

The magically-ideal steady-state income source for a private DARPA is an endowment that can fund operations from its interest. An endowment could create a near-infinite time horizon with minimal pandering to external stakeholders. Of course, getting to that point will take a long time and may be impossible!

Perpetual endowments have precedent for enabling different research paradigms

— the [Rockefeller Foundation](#), [Howard Hughes Medical Institute](#), and [Wellcome Trust](#) have all done great work.

What would the ideal steady-state budget of an DARPA-riff be? As some anchor points, DARPA's 2020 budget was \$3.6B¹⁹⁶ and Bell Labs' 1974 budget was weirdly similar at \$2.8B in 2021 dollars.¹⁹⁷ If it's true that research orgs don't scale, these historical anecdotes suggest that a single-digit billion dollar budget may be the upper limit for an effective research organization. This argument is admittedly hand-wavy argument but it provides a useful anchor for the high end of a long-term PARPA budget. Being aggressively ambitious, let's say ~\$1B / year. This number is less than DARPA and fully-evolved Bell Labs but more than Bell Labs' 1930 budget of \$300M in 2021 dollars. ¹⁹⁸

A back-of-the-envelope calculation tells us that to hit a \$1B budget, assuming 4% dividends on invested capital, you would need a \$25B endowment. That is an absurd, but not unthinkably large number. For comparison, Harvard, Yale, and Stanford have endowments of \$41B, \$31B, and \$29B respectively as of 2019¹⁹⁹. Howard Hughes Medical Institute had \$21.2B in assets in 2020²⁰⁰. Admittedly, these are the largest endowments in the world so the chances of being able to get there are incredibly low. However, they stand as proof-points that a long-term goal to build an endowment that could support a \$1B budget is hard but not impossible.²⁰¹

Even endowment-based funding is not without pitfalls. Endowment-based funding's boon and bane is that it dampens feedback loops. As a result the organization could become decoupled from its original goal as new generations

¹⁹⁶ See Department of Defense Fiscal Year (FY) 2020 Budget Estimates

¹⁹⁷ See The End of AT&T (might require a free account login.)

¹⁹⁸ See The making of American industrial research : science and business at GE and Bell 1876-1926

¹⁹⁹ See the list on wikipedia here

²⁰⁰ See HHMI Financials

²⁰¹ Contra an 'impossible' situation where the number came out bigger than the market capitalization of Amazon, for example.

of people (specially professional managers) take over. While this problem is not inevitable, the distinct possibility emphasizes the importance of long-term focused governance structures and intergenerational culture.

Another potential downside of loose feedback loops is that mission-focused organizations run the risk of perverse incentives when they come close to achieving their mission. Instead of closing up shop at a job well done, mission-focused organizations can wind up exaggerating the prevalence of a problem or slacking on their job in order to remain relevant. Luckily for a DARPA-riff, I'm not too worried about running out of possibly potentially game-changing but otherwise unsupportable projects.

Another minor consideration is that the other side of endowment-based funding's stability is that it can't quickly scale up. That doesn't seem like a significant limitation because we can expect private DARPAs to be more ideas limited than money limited.²⁰² Anecdotally, a DARPA director once asked for congress to *decrease* their budget, lest they be crushed under the accompanying expectations.

Of course, the biggest question is "how do you build an endowment in the first place?"

Let's rule out the possibility of acquiring the money all at once for three reasons. First is the obvious one — anybody with a chunk of money that big burning a hole in their pocket already has plans for it. Second, it is important for big things to start small — the organization needs the flexibility to work out kinks and develop a culture without the pressure to deploy a massive amount of money. Third, most people with the ability to create an endowment of that size would (reasonably!) want near direct control of the organization, which would likely as not lead to headaches.²⁰³

²⁰² See What makes DARPA tick

²⁰³ See this essay by Paul Graham about unrestricted donations

Ruling out one big funding event means that a private DARPA will need to build an endowment up over time, depending on the other funding sources we've explored in this section. Through a conscious process of budgetary discipline the organization could squirrel away funds, perhaps helped by a windfall here and there. Again, an endowment is a possibly unreachable goal, but it's one of those low-probability events whose probably drops to zero if you don't shoot for it.

There are some parts of the DARPA model that if changed, would make it hard to even call an organization a ‘DARPA-riff’ in the first place. Other experiments are more ‘optional’ — they might make the difference between success and failure or they might be a distraction. These ideas are kind of like clothes for an arctic explorer — they’re absolutely critical to success but are fairly interchangeable and not core to the explorer’s core identity or humanity. I’m going to present them in a way that feels epistemologically very certain; the alternative would be juicing every sentence with so many qualifiers and softening words that each paragraph would have three times as many words and one third the information content. So while the framing will be “Here is a thing. It will be good for this reason. It will be bad for that reason.” you should read it as “I have a hypothesis that this thing could work. I suspect that it might be good for this reason. My hunch is that it could be bad for that reason.”

Exploratory Program Organization

One potential experiment is that each program could be its own semi-independent entity - a Exploratory Program Organization (EPO). You could imagine it as a legal entity that enables a research program to smoothly slide between the fully externalized research that DARPA coordinates and the internalized research of an organization like [Lockheed Skunkworks](#).

DARPA doesn’t do any research in house. Externalized research has benefits, especially at the beginning of high-risk programs. It can act as a buffer against B-players, allows you to tap into tacit knowledge and specialized equipment without going all in, and more. Externalized research also has downsides that increase as the research matures. Dependencies between different parts of any research program tend to increase as the program matures, increasing the coordination costs if everybody involved isn’t under the same roof. Anybody who has done a group project has experienced this: it starts with “you go figure out how we’ll do path planning and I’ll figure out how to talk to the motor” but the division of labor breaks down when you actually need the motor to do the

path planning! These tradeoffs mean that there's no constant correct level of internal/external research throughout a program. It's even possible that one reason DARPA has a mixed record on transitioning is because their pure externalized research makes it hard to cohere the different projects into a single system.

At the beginning of a program, externalized research is important for building conviction. Externalized research enables the program to have access to rare skills or equipment that might only exist in one or two places in the world. Perhaps there is just one researcher who has perfected the technique for synthesizing a particular molecule. Externalized research allows the program to tap into people who would only want to work part time or have other commitments like finishing a PhD. One of the big barriers to starting a company is that founders need to be 100% in or there is weird murky uncertainty. Many DARPA programs would have been killed by these coordination costs if they had to be a self-contained organization from day one.

At the end of a program, internalized work is important for bringing disparate pieces of a program together and commercializing or otherwise 'encapsulating' a technology so other people can use it. While it's possible, there are very few examples of successful decentralized productization and someone needs to buy technology eventually.

In its early stages, an EPO could look similar to a DARPA program, with a program manager coordinating externalized research. As the program matures, performers can move from academic labs or other organizations to directly work for the EPO. This internalization lowers coordination costs around system integration once the individual pieces are de-risked. There's something clutch about the ability to go to someone's desk and say "hey look at this" when you're dealing with physical systems. In effect, the program manager is an initial CEO for the EPO, gradually internalizing outputs and people from the externally funded activity.

It is hard to know going into a program what kind of value it will create. Some

research programs end up creating something that resembles a product, some end up creating valuable IP, and some just end up creating valuable knowledge without a clear way to capture it. Of course many programs produce nothing of value at all except good stories. *The EPO acknowledges this reality by being temporary and converting to one of several options at the end of its life.*

At the end of it's life, the EPO could:

1. Renew for another life if several more years of external research could get the program to a much better place.
2. Convert into a profit-focused C-Corp or LLC if it has reached a point where a product is on the horizon.
3. Convert into a nonprofit focused on advancing the EPO's research mission if the EPO's research needs long-term research and coordination but won't be producing a product.
4. Convert into an organization for licensing IP if the EPO produced valuable IP that doesn't cohere into a product.
5. Or (in reality the most likely case) the EPO can do nothing in particular after it expires.

Many EPOs will not work out. That is ok because of the continuously varied commitment and temporary nature. You're not asking investors to put in a ton of money or people to quit their jobs from day one. The high failure rate might suggest that an EPO should be structured to be renewed every year with a maximum lifetime of five years. You wouldn't want a zombie EPO sucking resources for years after it's clear that it won't work.

EPOs could give PMs incentives that DARPA programs cannot — the ability to become the leader of an organization devoted to making a vision into a reality. In a way, the PM to EPO leader would systematize Robert Taylor's role in the personal computer; After running the Information Processing Technology Office at ARPA, Taylor spun up the interactive computing group at PARC and recruited

the people who had been working on the ARPA program to work there.

EPOs could bridge the valley of death. A big ‘failure’ mode of both DARPA and ARPA-E is that nobody takes the technology football when a program successfully wraps up. It’s especially insidious because that handoff falls into the cracks between job descriptions so often someone has to go above and beyond to make it happen. In the book [Loonshots](#), Safi Bahcall uses the pithy aphorism “Manage the Transfer not the Technology” to emphasize the point that this handoff, not the technological development itself, is where many promising technologies go to die. Although ARPA-E has a commercialization team, they only try to get the individual pieces of a program commercialized, but do not coordinate them into a single entity. The nature of an EPO removes that gap entirely; the program *becomes* a new organization that takes the football. A continuous slider from modularized external research to coordinated internalized research also enables people from disciplines farther down the ‘pipeline’ to be involved in the process. For example, prototypes are much more likely to be manufactured if people from manufacturing are in the room during the prototype design.

EPOs probably need a novel legal structure.

A parent organization for the EPOs

You could imagine EPOs acting as stand-alone organizations; something akin Focused Research Organizations²⁰⁴ that initially have a distributed structure. However, EPOs will be drastically more effective with an umbrella organization that can spin them up, and maintain things the nebulous things that benefit from institutional consistency like intra-organizational relationships and tacit knowledge. In other words, EPO’s are a complement to a DARPA-riff, not a replacement for it.

A DARPA-riff is important for getting EPOs off the ground. Starting EPOs

²⁰⁴ See Focused Research Organizations to Accelerate Science, Technology, and Medicine.

probably involves an involved program design process and scattered experiments to show that it's worthwhile to start. You need another organization to do that work. Additionally, DARPA PMs often work on multiple programs at once, which would be harder to do without a DARPA-riff acting as an umbrella organization. Culture, process knowledge, extra-organizational relationships, and organizational trust that a parent organization could build are intangible but could ultimately make each individual EPO more likely to succeed both in terms of know-how and money.

Since any given EPO is likely to fail, an umbrella organization is important for creating a portfolio of programs and shifting money between them. The portfolio approach enables funders to put money into the parent organization with a much higher chance that it will have a payoff than with any individual EPO. If each EPO needed to raise its own money, it would take a lot of time and resources and they would be incentivized to work on whatever is most likely to be funded, which heavily incentivizes likely-to-work or heavily hyped areas.

Issues

The private companies and labs who DARPA (and hence a DARPA-riff) works with might have a problem with folks leaving to join an EPO. "This might lead to people leaving" might make other organizations hesitant to work with you in the first place. Graduate students and post docs who are transient members of labs are the obvious exception.

Intellectual property ownership will also be an issue. Companies and universities are willing to work with DARPA at least in part because they know that they will get to keep the IP that comes out of the collaboration. At the same time, large companies work with universities and smaller companies all the time and presumably have satisfactory IP arrangements so this doesn't seem like an unsolvable problem.

You can create a program design discipline that enables better research and development

There is a lot of literature about how to manage research programs, but very little thinking about how to design them in the first place.

Program design is inextricably tied up with research planning. The idea of planning research might make some people throw up in their mouths a little bit²⁰⁵, but bear with me. I'm going to put on my Wittgenstein hat²⁰⁶ and suggest that words matter; what people are actually objecting to is *scheduling* research around *deliverables*, which is very different from *planning* research around *goals*.

All research is planned around goals to some extent. Even "I'm going to put this liquid in this other liquid and see what happens" is a plan. In this situation your goal is to see and report on what happens you mix those two liquids. It's a silly example but it provides a base case to build off of. The chemist mixing the chemicals probably has a bigger plan than simply seeing what happens; the plan probably involves mixing many combinations of chemicals in order to chase a bigger goal of eg. understanding which combination produces the most heat or perhaps why exothermic reactions exist at all.

Our chemist friend doesn't necessarily have or need a specific deliverable to do good work. It might literally be impossible for her to, for example, find a mixture that produces as much heat as she wants. Even if she could, it might involve testing an order of magnitude more mixtures or even needing to invent an entirely new synthesis technique to do it. Requiring her to work on a specific schedule would be inane. I suspect that deliverables and schedules are artifacts of the commoditization process that happened to research during the 20th

²⁰⁵ See Fund people, not projects I: The HHMI and the NIH Director's Pioneer Award and the other parts of the same series for a deep dive on the literature around this.

²⁰⁶ See Farnam Street — Wittgenstein: Reality is shaped by the words we use

century as a byproduct of the expansion and institutionalization of science.²⁰⁷

If planning and goal setting (separate from schedules and deliverables) are necessary for any kind of R&D work, it raises several questions. What kind of goals should be set and what timescales should be planned for what sort of work? Should you parallelize work or serialize it? In what contexts? What are the key constraints keeping you from a breakthrough? What level of fidelity will stifle creativity and what will provide the focus to make breakthroughs? Which tools will maximize the focus and minimize stifling and which frameworks can make the difference between good work and bad? I would posit that these questions and their ilk, what I might refer to as ‘program design’, are woefully under-theorized.

In addition to the big-headed questions, the objective truth seems to be that we kind of suck at designing research-y programs. Most programs feel like a grab-bag of vaguely related projects that are incremental extensions of whatever researchers were already working on but nudged slightly towards a central theme. Most technology roadmaps feel like the penultimate committee output; bland documents seem to both say everything but are useless for making hard decisions.²⁰⁸ Whether or not we’ve actually become *worse* at program design is an open question. To stake a firm position, I will say yes, we have become worse — anecdotally compare F-35 or SLS development to Apollo. This disciplinary decay would due to reasonable incentives — there are fewer areas that need program design to function. Arguably software has become the dominant ‘innovation-based industry’ and, unlike science-based manufacturing, it doesn’t benefit significantly from (and can even be harmed by) long-time-horizon program design.

The question of whether we’ve become worse at program design is an important

²⁰⁷ See The decline of unfettered research for a good description of this process and its consequences.

²⁰⁸ There are some exceptions! The International Technology Roadmap for Semiconductors is quite good, I presume because an entire massive industry *actually* depends on coordinated research.

preliminary for the more important question: “could we do program design better?”²⁰⁹ If we’ve gotten worse, it’s strong evidence that we could do better. However, I suspect that we can do better regardless of whether we’ve always sucked at it (or are secretly good at it? 🤔)

One reason program design is under-theorized is that the people who are good at it are too busy applying it to spend time and effort making it legible.²¹⁰ There are no research coaches. Program design also suffers acutely from siloization — each organization (or individual!) seems to reinvent the wheel. This isn’t to say that we should aim for standardization; program design will always be very context-dependent. However, it seems possible to shoot for a baseline of best practices and frameworks for building on top of those best practices.

There are many inspirations that the discipline might be able to draw on. A big part of a program design discipline will be trying to make legible the process that skilled practitioners already use to create things like [Adam Marblestone’s Climate Technology Primer](#) or [José Luis Ricón’s Longevity FAQ](#). José does a good job unpacking his process in [Longevity FAQ: Making Of](#); one could imagine a line of research that focuses on understanding why some people are good at this process and some are not. Adam and Ed Boyden hinted at the possibility of more structured discovery in their article [Architecting Discovery](#).

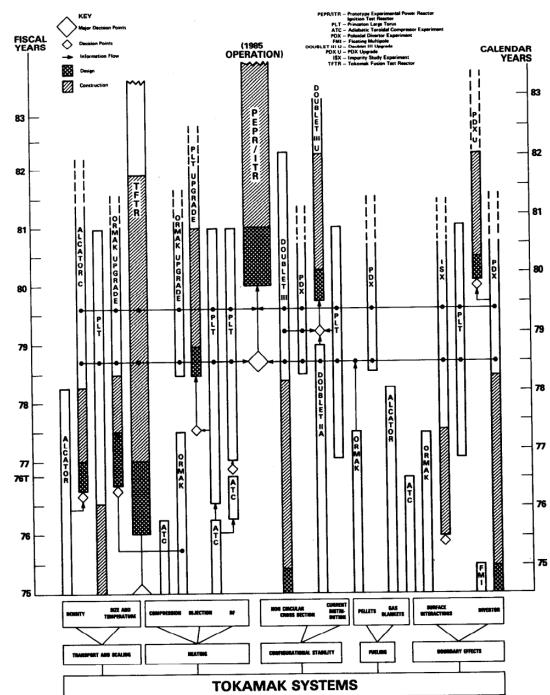
History is another place to dig — how did people manage research-adjacent projects that had [shockingly fast results](#)? While some of the results were contingent on the individuals involved, I can’t shake the hunch that there is transferrable knowledge in the tools developed to manage early space programs.²¹¹ How did the creators of the [Fusion Power By Magnetic](#)

²⁰⁹ Yes “better” is a vague word. In the context of program design means some combination of “faster, cheaper, more successfully, enabling programs that wouldn’t exist otherwise, and generating more knowledge in both successes and failures.”

²¹⁰ Many disciplines which are meant to supplement practice suffer the same fate!

²¹¹ See *The Secrets of Apollo* or the PERT guide for Management.

[Confinement report](#) think about the different possibilities and tradeoffs?²¹² Of course, the value of historical study depends on the aforementioned question of whether we've become worse or not.



You don't often see this kind of program design any more. Was it never useful, now outdated, or a lost art?

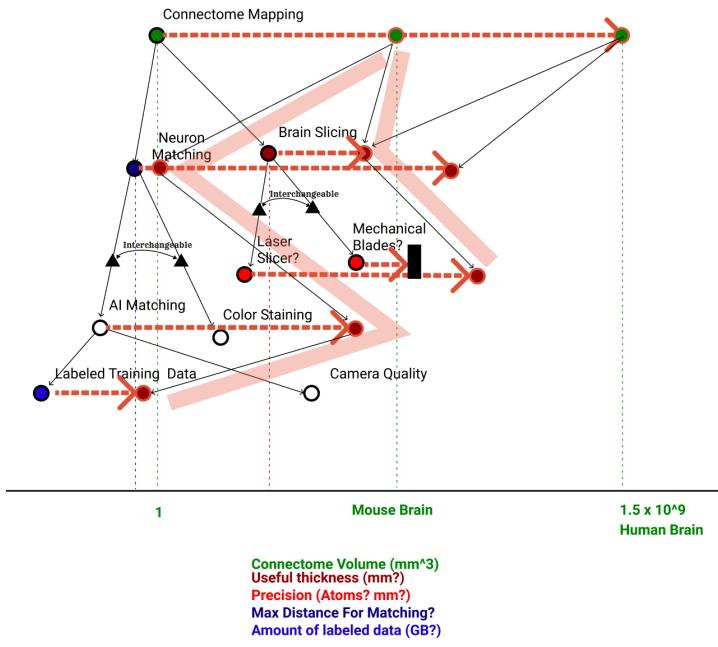
There has been a lot of work on how to do better project (as opposed to program) management, which is incredibly relevant. One could even argue that researchy program design is just part of project management. I would argue that standard project management techniques (Six Sigma²¹³, Matrix Management²¹⁴, etc.) need so much modification to deal with the uncertainty, timescales, and fractured expertise inherent in research programs that we're talking about something beyond a simple extension of project management. There is a whole range of

²¹² I have tried to track them down to ask them but they seem to all be unreachable or dead. If you know any of them, please put us in touch!

²¹³ Six Sigma is Motorola's process-improvement process.

²¹⁴ Matrix management is a now-disfavored project management approach created by the aerospace industry in the 1950s.

methodologies like TRIZ²¹⁵, [Wigmore Charts](#), or [Wardley Maps](#) that feel like they have nuggets of truth but are, for the most part, post-hoc explanations of success used to get consulting gigs in areas with loose feedback loops like big companies or law schools. So, like any good discipline, program design would start strongly adjacent to crackpots and mysticism!²¹⁶



What might a Wardley-inspired roadmap for describing the complete structural connectivity of the human brain look like?

There are several areas that I suspect are important for good program design where I haven't found much written prior art (unless you squint *very* hard): Systematic ways to think about technological constraints and tradeoffs; Ways to visually represent constraints and possibilities in a way that can actually aid decision making and reveal new possibilities; and the tactical psychology of creating programs. Each of these areas is something that people already do but in an ad-hoc way. Any time someone says something along the lines of "oh, you

²¹⁵ See And Suddenly the Inventor Appeared: Triz, the Theory of Inventive Problem Solving

²¹⁶ See: alchemy, astrology, humor-based medicine.

can't do long-range battery-powered commercial flight — batteries are too heavy" they're implicitly bundling many constraints and assumptions: What kind of batteries are we talking about? What is their power density? What components contribute what fraction to their weight? How much power do they need to put out for how long to be useful? What are we assuming about the propulsion system for the plane? What are the degrees of freedom for all of these these components? What are the dependencies between them? How hard would it be to improve any one of those components by how much and what would that effect be on the whole system? The same sort of questions are important for any technology. That many complex questions are also impossible to hold in your head all at once, let alone transfer into someone else's head — thus the need for thinking frameworks and representation tools.

And then there's the shockingly human and under-discussed process of mining that knowledge in the first place. Perhaps controversially, I'm convinced that most human knowledge is not encoded, especially on the knowledge frontier. As a consequence, a lot of program design is actually applied psychology. As the program designer you need to not only nudge people to talk about their area of expertise in a way that they probably haven't done before and generate excitement about the idea of a bigger program, you need to figure out *who* are the right people to talk to and get them to talk to you in the first place! This piece of program design resembles some combination of sales and user research. Equally (or perhaps more!) important to the question of "what work needs to be done?" is "who is best suited to do the work?" It's quite funny actually — while there seems to be a strong cultural consensus that good outcomes in high-uncertainty endeavors depend heavily on the specific individuals involved, answers to the question "who is best suited to do this work?" Are almost entirely absent when people lay out a research program.

It's important to acknowledge that this entire hypothesis that you can create a program design discipline could easily be false. Perhaps guiding research is more like creating art; one can talk about specific techniques (**art**: stenciling, mixing paints, shadows | **program design**: talking to experts, budgeting, evaluating

claims) but the process itself is too context-dependent and holistic to improve through the tools and abstractions that a discipline would provide. Perhaps more dangerous is the possibility that any formalization ventures out of the realm of planning+goals and into the realm of schedules+deliverables, hamstringing the work you intend to enable.

Another strong argument against the formalization of program design is that the real issues are Rumsfeldian unknown unknowns that are undiscoverable or unthinkable before you run into them in the process of doing the work. Paradigm shifts are to some extent unimaginable before they happen by definition. The presumption that we could do a better job than practitioner intuition could be a foolish waste of brain-cycles at best, or actively steer people away from actually useful work at worst. “How do you draw a roadmap off the edge of the map?” Is a valid question.

The possible upsides of a healthy program design discipline seem like they could be huge. So if you go in explicitly acknowledging the potential failure modes, the potential upsides are worth it. Program design could potentially help programs exist that would not exist otherwise, prevent wild goose chases, and unlock unintuitive possibilities!

The trick is that to do risky experiments²¹⁷, program design as a discipline needs to be embedded in an organization running programs. Enabling technologies must be developed while doing serious work²¹⁸ and without an organizational home, program design would be severed from a serious context of use. It would suffer the same fate as other academic or consultant-driven disciplines of practice. These disciplines tend to sound great on paper so that someone wants to implement them, but that implementation often has so many restrictions, imperfect information, and strung-out feedback loops that it’s unclear how much of the success or failure of the project is attributable to the discipline. Developing better program design is one of the reasons why despite running multiple

²¹⁷ In the Popperian sense.

²¹⁸ See How can we develop transformative tools for thought?

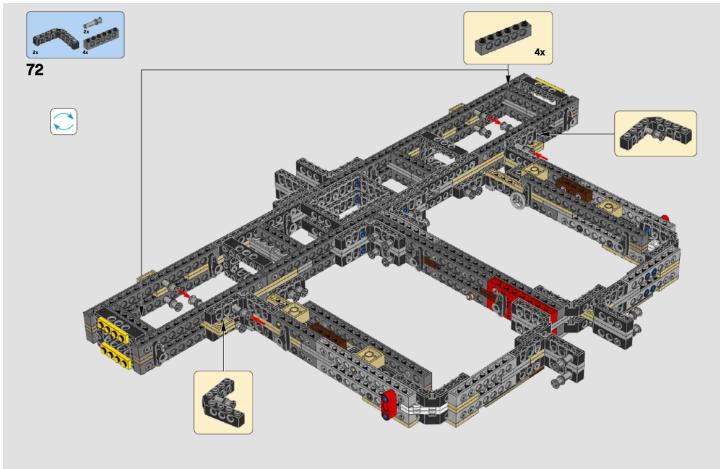
organizational experiments flies in the face of common wisdom, a DARPA-riff both can and must do multiple experiments at once.

Program design is like creating the lego instruction book

'Program design' is an awfully nebulous term. Its centrality to any DARPA-riff makes it worth spending some time wrapping our heads around. Like many nebulous things, the best way to think about program design may be through an extended analogy. If you didn't build Lego sets as a kid I recommend you pause reading, buy one, build it, and then resume.

Modern research institutions tend to produce a lot of atomic contributions. These contributions resemble lego bricks in the sense that you can intuit that while they are not useful on their own, they will be useful as part of a broader system. But like most lego bricks it's not immediately clear what amazing thing the brick can eventually be part of (and in fact it can be part of many things!) Even a giant bucket of lego bricks often doesn't immediately suggest what to build. This 'bucket-of-bricks' is the state of many disciplines. That description is too extreme — the situation is more that we have the picture on the front of the box and the bricks that came in the box mixed with a bunch of extraneous ones. Thought leaders and big ideas people talk about the different box-pictures all the time, everybody has a vague idea that their brick is a part of the spaceship. But if you've ever put together a lego set, it's rarely clear where each part fits in until you see all the intermediate steps in the instruction book.

The goal of program design is to create these intermediate steps. For example, the painfully unsexy step 72 that has a bunch of tan and grey pieces in a rough shape with some flanges on it that doesn't look like anything at all. It is *hard* to know that step 72 is going to lead to the awesome spaceship on the box; it certainly doesn't look like a spaceship.



It's a spaceship, obviously.

Getting step 34 to happen is hard. You need to imagine the unseen interior of what's on the box, which pieces are available, which pieces you would ideally have, and which substitutions are ok (it doesn't really matter if you use a neon green brick instead of a grey one as an interior part). If you really need a piece that doesn't exist, you need to make it yourself or convince someone else to make it. And then you need to get all the pieces together — just because you put together an actionable plan does not mean people will take those actions!

Ideally, as many steps as possible are relatively self-contained ‘modules’: the cockpit part of the spaceship, a secondary landing craft, an escape pod. In real life, this corresponds to intermediate goals that are useful on their own or are at least exciting milestones or demos. There are many words that should be written about the tension between the benefits of useful intermediate steps and the danger that undue focus on useful intermediate steps can pose to eventual outcomes. In hindsight it’s easy to both look at technological development and forget how many incremental useful steps there were (synthetic plastic was used as a nifty liquid by printers before it was accidentally cooked) and to forget how much work has gone into technologies before they get to *anything* useful (transistors needed eight years of work before the first completely unusable point-contact demonstration).

Obviously, this analogy can only be pushed so far: research-heavy programs will

inevitably have gaps between steps that would make any lego-assembler tear her hair out; the picture on the front of the box might be fuzzy or wrong; in real life, pieces don't just snap together; technology doesn't exist independent of people, so you need to deal with incentives, politics, and egos; and more.

You could reasonably object to this analogy with a counter-analogy that the pieces of a research program are less like legos and more like biological components. Separate components in a biological process have much more 'agency' — if you create the right environment and shake it or apply a jolt of electricity, they will self-assemble²¹⁹. The biological analogy would be much more apt than a lego analogy if paradigm-shifting work cannot be planned or the innovation ecosystem is 'directionally' correct and just needs more funding and work along existing channels.

I suspect the reality is a mix of both analogies — in research programs, unlike lego construction, you can't lay out every single step so 'biological' self-assembly between steps will be essential. Keep in mind that these analogies are meant as thinking tools, not evidence or predictive descriptions of the world. Hopefully, though, the Lego analogy is provocative. Playing with it is a more tangible way to describe the need for program design overall and suggests extensions, like the the possible role of simulations and legible intermediate steps.

Good Simulations could be the 'why now' of roadmapping tools

New technologies and disciplines are often 'good ideas' in theory but infeasible in practice because they're built on top of other technologies that aren't up to snuff. Electric cars that are comparable to gasoline-powered cars weren't really feasible until the electronics industry caused leaps in lithium-ion batteries. "Why now?" a good question to ask whenever someone proposes a new technology or discipline. A good answer can hint at eventual success. Of course, the many

²¹⁹ See the Miller–Urey experiment.

examples of successful creations that didn't have a particular "why now" means that while a bad answer to the question doesn't predict failure. Simulations may be an answer to the "why now" question for a new program design discipline: our capability to simulate everything from proteins to cities has exploded in the 21st century and they have unique ability to precisely illustrate possibilities that do not yet exist in a way that is both internally consistent and externally consistent.

Simulations have continuously improved over the late 20th century and early 21st century thanks to improvements in both software and hardware.²²⁰ Games and movies have pushed the edge of the possible, encouraging better algorithms, processes, and tools. Improvement in Graphics Processing Units (GPUs) both directly enable better simulations (the world is massively parallel!) and have fueled improvements in machine learning which has recently started unlocking simulation capabilities like AlphaFold and FermiNet.²²¹

With simulations, you can *precisely* work backwards from a long-term goal technology. This "backwards propagation" can then be used to inform what you build starting with what is possible today. In a way, this role is similar the one fulfilled by science fiction when you use it as a case study, or simply a precise vision of where a technology can go. However, simulations are much more precise than visions and can be not only a compass but a map as well. They can surface obstacles and make non-intuitive suggestions about which routes might be productive and which might be traps.

Take 'atomically precise manufacturing', for example. There is one camp that says "there's no reason it shouldn't work" but doesn't give clear examples of what "working" would actually look like. There is another camp that says "that's

²²⁰ Simulations, specifically simulating nuclear detonations, were among the first applications of digital computers.

²²¹ See AlphaFold: Using AI for scientific discovery and FermiNet: Quantum Physics and Chemistry from First Principles

not how molecules work!” but like the proponents, can’t be any more grounded. With that level of discourse the two groups inevitably just talk past one another which makes it hard for even a technically trained outsider to evaluate their arguments. In a way it resembles the post-consensus political discourse. If both groups could agree to the assumptions and methods behind a simulation, it would be possible to create more-or-less objective evidence for or against the feasibility of one plan or another. The simulations could then enable researchers²²² to iterate on precise challenges.

Even more optimistically, simulations could enable non-researchers to contribute as well! If you squint, video games are simply simulations optimized for fun instead of accuracy. One could imagine many <games? Tools?> somewhere between [Foldit](#) and [Kerbal Space Program](#) that enable non-experts to explore possibility space and surface approaches orthogonal to expert thinking.

Of course there are challenges. To name a few: it is hard to interface different simulations; it is hard to know where a simulation diverges from reality; and simulations and models often fail to capture important aspects of a real system. Like any intellectual tool, simulations can be used well and poorly. People often claim that an idea works ‘in simulation’ without laying out the assumptions behind that simulation and without any connection to how it might be physically implemented. So this isn’t to say that simulations can address every problem right now. However, the space of problems that simulations can potentially unlock has significantly expanded in the past several years (as of 2021).

A DARPA-riff could use a constitution with checks and balances to produce high variance results over a long period of time

The DARPA model is an unbuffered system. In chemistry, an unbuffered system changes its overall pH rapidly in response to acids or bases. At DARPA,

²²² And ideally non-researchers! See: Foldit below.

individual contributions easily have a huge effect on outcomes. An unbuffered system has many advantages — most notably it enables high agency individuals to act quickly with minimal friction. At the same time, an unbuffered system seems likely to run into a catastrophic failure that kills the whole thing. In order to do effective solutions R&D, a DARPA-riff needs to be a long-term institution. Is it possible to build in checks and balances that could sustain a long-term institution without ruining the possibility of high variance outputs?

Looking far outside the realm of research organizations, it's worth noting that there are some core elements to the American constitution that have made the United States a rather robust institution.²²³ The US founders thought a lot about institutional longevity and it's worth drawing from their experience. As a thought experiment, let's go back to 6th grade civics class (for those readers who didn't go to school in the US, you really didn't miss much). We can examine the roles of different branches of government and play with them through the lens of the different building blocks we have in corporate governance and DARPA's structure. I'll pay special attention to the Buxton Index (the length of the period over which an entity makes its plans) of each role to explicitly call out where there will be incentive mismatches. Ideally, you can use those tensions as a productive set of checks and balances instead of generating wasteful friction.

Different Roles as Different Branches of Government

It seems pretty clear that the program managers (PMs), as the people who are actually executing on the programs, are analogous to the *Executive Branch*. Similar to the American executive branch, they probably have the lowest Buxton index. Remember, PMs have a tenure of four to five years. PMs are also the outward facing piece of the institution, similar to the role of the executive branch.

The role of the *Judicial Branch* seems well served by some kind of multi-person board of directors (which is an awkward juxtaposition with the "Director" titles usually held by DARPA managers so we'll just call it the Board for now). Like the Judicial branch it makes sense for the board to be the farthest removed from day-

²²³ I would argue that this is the case despite recent (2021) events.

to-day operations of the organization. Also similar to the Judicial branch (and other boards) it seems like a good starting place to have it be composed of several people with indefinite term limits, both giving it the highest Buxton Index and the lowest variance; we can use the fact that decisions made by committees lead to median results in our favor.

That leaves the Director(s) as the *Legislative Branch*. At first this connection seems a bit odd because we're juxtaposing a bicameral legislature and at most a few people (and likely just one person at least at first). However, taking a step back and looking at the Legislative branch as the internally facing branch that writes the laws, the connection makes more sense. The Director will have a Buxton Index that is somewhere between the PMs and the Board and be responsible for setting culture and procedures, money, and other internal functions that are roughly analogous to laws.

The last piece of the puzzle is a constitution and a body of laws. Most corporations have a constitution but it's usually a formality. Instead, a DARPA-riff should take a constitution seriously as a coordination mechanism. It flies in the face of common wisdom to have written rules early on in an organization's life. However, as a mechanism for enabling institutional longevity it might be valuable enough to introduce it as a source of friction. Many fast-moving organizations still have very writing-centric cultures. You could take this approach one step further and say that memos constitute a body of laws.

Checks and Balances Between Branches

The checks and balances should be based around two principles: 1) by default people should be allowed to do things unless they will harm the organization's long-term ability to achieve its goals 2) each branch should be able to check each other branch.

It's standard for a board to appoint CEOs (in this case, directors). This practice fits into our framework. Since there is no electorate, the board is effectively the representative of stakeholders who would otherwise elect the members of the legislature. Similar to the judiciary, the board should be able to step in and stop

actions that *explicitly* go against the organization's written constitution. This less-standard ability gives them a lot of power to check both the Executive and Legislative branch, but they are only allowed to use it in a very constrained way.

In most corporations, the board has the ability to remove the CEO. Should that be the case in a DARPA-riff? It gives the Board a ton of power to warp the director's incentives. My hunch is that it should be possible, but should resemble impeachment in the sense that it needs to be an action of last resort and needs to be non-arbitrary (ie. Constitutionally based).

Like the American legislative branch's check on the executive, money will be the Director's major check on the PMs. For this to work, trust and communication has to flow freely in both directions. The director's check on the board is through the written rules, as the board is constrained to only stop actions that go against the written rules. Giving the director an explicit check on the board is one place where we deviate from normal corporate structure.

The PMs/Executive Branch will require some non-standard tools in order to have official checks on the other two branches. Normally employees don't have an official check outside of perhaps choosing to join or leave an organization. Although their positions are explicitly temporary, you want PMs to be as bought-in to the institution as everybody else. Perhaps even more so. Having their own levers of power are important to that buy-in.

Given that PMs should have checks on the directors and the board, what should they look like? In the US government the executive branch nominates new justices. PMs nominating new board members is unconventional but it creates an intriguing coupling between the lowest and highest Buxton Indexes in the organization. The hope would be that the PMs would nominate people who would *get it*, in the sense that it takes someone who has been in the trenches to *get* what it's like to be there. Board member election could also be a mechanism to get long-term PM buy-in despite short-term tenures. Either 'graduated' PMs could participate in every board member selection or there could be a single board position devoted to representing alumni that only alumni can select. More

broadly, employee-ownership could enable a DARPA-riff to align long-term incentives.

The PMs' check on the Director could also resemble the American system, where the executive branch has veto power over the legislative branch. Perhaps the PMs should have veto power over both proposed rule changes and (especially) new PMs. Implementing veto power within a group of people is tricky because it could easily lead to politics with single PMs trying to convince the others to go along with them.

Avoiding Deadlock

Deadlock is the primary failure mode of an organization with a robust system of checks and balances. The obvious places in this scheme that seem most deadlock-prone are the PMs' veto power and the writing-based decision making. You could imagine a situation where the PMs veto literally every rule change and potential new PM to try to force the Director to do something. Similarly, you could also imagine a situation where everybody uses written precedent and the constitution to browbeat and block each other instead of talking and trusting.

To the former case, there could be a check on this check like enabling an agreement between the director and the board to overturn a veto. To the later case, one could suggest something like a “ask forgiveness not permission clause” that enables rule-breaking actions if they have good outcomes. Both of these might be fine ideas and there may be other similar fixes, but there are limitations to this type of thinking.

At the end of the day, you cannot create a set of rules that will *guarantee* that a group of people will work together in a way that produces high-variance results over many organizational generations. Effective action will always depend on trust and communication. There are several structural pieces that can make this trust and communication *more likely*: keeping the organization small; maintaining a precise mission; and filling all positions with excellent, diligent, and applied people. However, all structural components will be worthless if the people in the organization don't commit to constantly maintaining the organization in the

same way you maintain a building. Intentionality and trust are unavoidable!

Employee-ownership could enable an ARPA-riff to align long-term incentives

A DARPA-riff needs to be obsessed with program managers. Unfortunately, many of the attributes that make people excellent PMs also make them good for many other roles; people who are potential PMs have a lot of options. Unlike DARPA, a DARPA-riff doesn't offer people a way to serve their country and at least in the early days, you won't be able to pay as much as a tech or finance job.

This heavy reliance on in-demand people means that it's important to think seriously about why people would take up the role and how to make sure that their incentives are aligned with the organization's. There are many ways a DARPA-riff could structurally address this question; 'employee ownership' might go a long way towards encouraging everybody involved in the enterprise to work towards its long-term success. Of course, that term is incredibly vague so let's break it down and think about what implementing it might look like.

Employee ownership via economic stake

One sense of employee ownership is to give everybody a stake in any potential economic upside. While a DARPA-riff probably can't be profitable *overall*, that doesn't necessarily mean there can't be economic upside for employees.

It's likely that a DARPA-riff will eventually spin out companies. If it does, it could potentially retain shares in those companies via a for-profit entity. Employees could have shares of this entity. The shares could act as a smoothing function on the amount of value different PMs capture and create buy-in to the organization's long term success instead of only a PM's specific program. Anecdotally, even a small number of shares in Entrepreneur First's global fund (which rationally most people realized was not going to be a life-changing

amount of money) created a sense of caring about the organization and its long-term outcome.

The best way to handle liquidity in a for-profit DARPA-riff subsidiary is still an open question. If employees have shares in a for-profit subsidiary, somehow those shares need to turn into cash eventually. The subsidiary could go public, but that's a bad idea for too many reasons to get into here. Another option would be to have it be a fixed-life fund similar to a PE or VC fund — this seems create misaligned incentives with creating long term value. Both traditional options being off the table means that the situation isn't hopeless, but it does need more work.

The first requirement that Nonprofits must meet to maintain 501(3)(c) Status might clash with employee an employee pool: "Organizations that apply for tax-exempt status cannot serve the private interests, or private benefit, of any individual or organization besides itself past an insubstantial degree." This requirement is probably why OpenAI is structured so that all the employees are employees of the for-profit subsidiary. A DARPA-riff could do something similar, with the for-profit subsidiary employing almost everybody and contracting them to a Nonprofit entity.

Regardless of how it's done, it's important that any employee value capture mechanism ignores how much value that employee themselves captured. Specifically, I'm worried about incentives that push people to work on programs that they think can create more-certain capturable value instead of less-certain (but potentially massive) uncapturable value.

There will be unavoidable inequality between the amount of value different PMs capture. Some programs will probably lead to successful companies; the people who do the work to get the company off the ground will inevitably reap the lions share of the rewards. There is so much more to starting a successful company than the technology it's built around, so it seems reasonable that significant chunk of its value should be captured by people who build the company instead of just doing the research. Some programs will yield complete duds. Some PMs

will do work that leads to highly-paid consulting gigs or nifty jobs. Some programs will create massive value that is impossible to capture. It's impossible for an employee value capture mechanism to eliminate these inequalities (nor should it) but it should attempt to *buffer* them.

My model of monetary compensation is that for most people²²⁴ it's not a strict matter of "more is better" but that there is a threshold that feels "fair" above which other incentives matter much more.

Employee ownership via control

In addition to economic stakes, people tend to feel more invested in the long-term success of an institution if they have a voice in its direction.

The most prominent way that organizations implement employee ownership is through the co-op model.²²⁵ As gross oversimplification: instead of a co-op having a management team who answers to a board, co-op employees *are* the management team and elect representatives who liaise with the board. There's something emotionally appealing in the idea of a private DARPA as a 'research collective.'

A DARPA-riff likely can't adopt a co-op model full-cloth; co-ops tend to be organizations sell concrete outputs on relatively short timescales like farms and retail stores. However, the idea of PMs comprising the executive branch in a constitutional system is shockingly similar to a co-op. In both cases the 'members' select who will be on the board. The difference is that the constitutional model has three separate branches while co-ops have two. I suspect this third branch (the director/legislative branch) is important for dealing with longer timescales and more nebulosity than normal co-ops.

A bit of weirdness for control-type ownership pops up because of PMs' short

²²⁴ There are of course exceptions and those people go work in finance.

²²⁵ I honestly could not find any great resources on how co-ops work. The best I could do is this Ohio State University site. If you know of any better ones, please let me know!

tenures.

If PMs don't feel a long-term connection with the DARPA-riff their level of autonomy will easily allow them to act in ways that are good for them but bad for the organization. This could look like taking on lower-risk programs so they're likely to have a win under their belt that they can leverage later in their career; or (if the PMs have strong control-type ownership) pushing raises that would give the organization far less runway. There are many actions that not overtly bad like embezzling, but if they were widespread it would quickly bring the organization its knees. Short tenures have the potential to amplify this sort of problem.

Universities are a good case study in creating a connection between an institution and short-tenured individuals. Despite being primarily composed of students who will be there for ~5 years (with high variance), universities both manage to maintain a stable intergenerational culture and convince alumni to care about the institution's long-term health after they graduate. This example makes it worthwhile to ask "how are Universities so institutionally and culturally stable?" One possibility to create long-term buy in is to have a board member who is explicitly a representative for 'graduated' PMs, in the same way that universities often have an alumni representative as part of their governance.

A DARPA-Riff could establish sales channels for frontier tech

There is nothing more difficult to take in hand, more perilous to conduct or more uncertain in its success, than to take the lead in the new order of things; because the innovator has for enemies all those who have done well under the old conditions, and lukewarm defenders in those who may do well under the new.

— Machiavelli, *The Prince*

Another important organizational experiment is to figure out better ways to get

new types of technology out into the world.

Frontier technology has bespoke sales channels. People trying to get new technology out into the world constantly need to do some combination of shoving it into ill-fitting but established channels (“We’re going to sell this jetpack like SaaS software!”) or creating a new way of selling from scratch. The latter introduces a whole new failure point for the technology and sometimes requires as much inventiveness as creating the technology itself. Making this process less harrowing would have massive impact.

Speculatively, a DARPA-riff may be able to address new technology’s problematic dispersion-related uncertainty through institutional knowledge accumulation, creditability, and ongoing relationships.

Knowledge accumulation

Over the years the startup world has built up cultural wisdom about how to sell software. Could the same thing happen for the wildly different technologies that might come out of DARPA-like programs? A general purpose telerobotic platform is incredibly different from a molecular factory and they both look nothing like a muon-catalyzed fusion prototype.

Despite massive technology differences, I suspect that a DARPA-riff could accumulate institutional pattern matching for desperate technologies if they explicitly set out to do it. This hunch is based in large part on personal experience trying to help dozens of ‘deep tech’ startups off the ground. There were hints of replicable patterns in building out successful sales channels in everything from hydrogen filters to grocery-store robots. Truthfully, I didn’t spend enough time doing it to pull on that thread and investigate where it went.

Institutional creditability.

Creditability matters in technology diffusion. People (and thus organizations too) are more willing to at least try something they’re dubious of if it has a name they trust behind it. Of course, it’s critical that the technology work well! But creditability can be that initial foot in the door. In the same way that individual

academic labs can gain credibility from the university as a whole, individual programs and people could lean on a DARPA-riff's established organizational reputation.

Ongoing relationships.

Technology diffusion inevitably depends on many other organizations: manufacturers, regulators, resellers, advertisers ... the list goes on. And people like to work with people they've worked with before. One big reason people tend to bet on industry insiders launching new ventures is because of the assumption that their existing relationships will reduce friction and make them more likely to succeed. At the same time, new technological paradigms often come from outsiders. What might look like a nefarious plot to suppress change²²⁶ can also be explained by outsiders not having the right relationships.

It might be possible to circumvent this problem by maintaining relationships with people in a wide array of other organizations at the umbrella-organization level. The alternative would be to force each program to develop new relationships or finding someone who has them on top of technical work. Even more speculatively, these relationships could be codified in actual contracts where external organizations (manufacturers or distributors) are part of a consortium that gets first crack at program outputs, similar to the MIT Media Lab's model.

This is one of those ideas that can sound good on paper but will be incredibly hard in practice. [ARPA-E has an entire tech-to-market team](#) and still has a mixed track record.

R+D Incentivization SPAC

I'm allowed to have one part of the speculation section that's just pure wild

²²⁶ Which does happen sometimes!

speculation, right? Ok here goes: It might be possible to create a SPAC²²⁷ that is designed to incentivize getting technology past a goalpost.

Normally SPACs either have a fairly broad mandate (“find a high-potential tech company”) or are created in order to take a specific company public. What if instead the SPACs mandate was to go after a specific set of technological milestones? Say for example, the ability to create an arbitrary chemical reaction at a specific site faster than a given rate with so many nanometer precision (which would be broadly, the specs for a company on the way towards atomically precise manufacturing).

In effect, an R&D Incentiviation SPAC (RDISPAC just rolls off the tongue right?) could act like a non-government advanced market commitment for a technology. Advanced market commitments are the government saying in effect “if you create something that meets these specifications, we guarantee we will buy N of them for \$X apiece.”²²⁸ The most famous advanced market commitment in recent memory is the US government’s approach to COVID-19 vaccines. Of course an RDISPAC would differ from an advanced market commitment because investment doesn’t automatically translate to customers the way an advanced market commitment does.

The milestones that a SPAC is looking for would need to be very carefully considered. The best technologies don’t always win, so there needs to be a strong argument that technical capabilities will translate into a long-term profitable business. Drug companies get around this problem because it’s basically guaranteed that if you create a drug that gets FDA approval for a certain set of conditions, insurance companies will pay for it. However in any other case, new technology has bespoke sales channels. Instead of a fairly legible process, creating sales channels for something nobody has seen before requires almost-magical persuasion. One of the reasons why VCs (accurately) care so much about

²²⁷ In a grossly oversimplified nutshell, SPACs (Special Purpose Acquisition Companies) are publicly listed shell companies that then merge with a private company, effectively taking it public without an IPO. Wikipedia article [here](#).

²²⁸ Anton Howes has a good Twitter thread about Advanced Market Commitments [here](#)

a startup's founding team is because much of a company's success has nothing to do with the technology and everything to do with sales and market creation.²²⁹ It seems like an open question how to both argue that a technology-based SPAC is a good investment and have its acquisition conditions be based only on technical milestones. Additionally, the cash injection from the SPAC needs to be able to get the technology to a place where you're seeing fairly continuous business improvement. If the SPAC acquires a technology company and the share price falls, it's unlikely that you'll be able to raise any more money.

If we talk about it in terms of technology readiness levels²³⁰ (TRLs), an RDISPAC could target either 3 (proof-of-concept), 6 (prototype in a relevant environment), or 8 (production version in a real environment). If the RDISPAC were targeting TRL 3 or 6, the technology would actually skip being a VC funded startup. If the RDISPAC targeted TRL 8, you would still need to raise private funds and build a business but the existence of the SPAC (ie. A clear exit option) would hopefully make it easier to raise funds.

RDISPACs could help create a healthier technology creation ecosystem. The organization that is doing the work to get technology projects off the ground and help fund them initially (ie. A DARPA-riff) could manage the SPAC well. We've already seen a few technology-focused VC firms like Lux Capital go down this route. The advantage of a DARPA-riff helping manage a SPAC would be that management fees could go directly towards creating more technology. Running both the beginning and end of the pipeline may enable a 'closed loop' technology development system, where the organization helping germinate the technology can capture some of its value without taking deadweight equity or licensing fees.

Of course, the closed loop scheme could potentially incentivize fraud! It's important to think about fraud up front because in addition to being bad in and of itself, it could poison the whole mechanism even for honest actors. Precise, *externally verifiable* acquisition conditions could both prevent fraud and create

²²⁹ See Productive Uncertainty

²³⁰ See a more detailed description of technology readiness levels here

alignment between investors and organizations shooting for the target. In a way, precise acquisition conditions can focus the conversation around how valuable it could be if technology with a certain capability existed. It's entirely my bias, but these conversations would further increase the value of precise, understandable roadmaps.

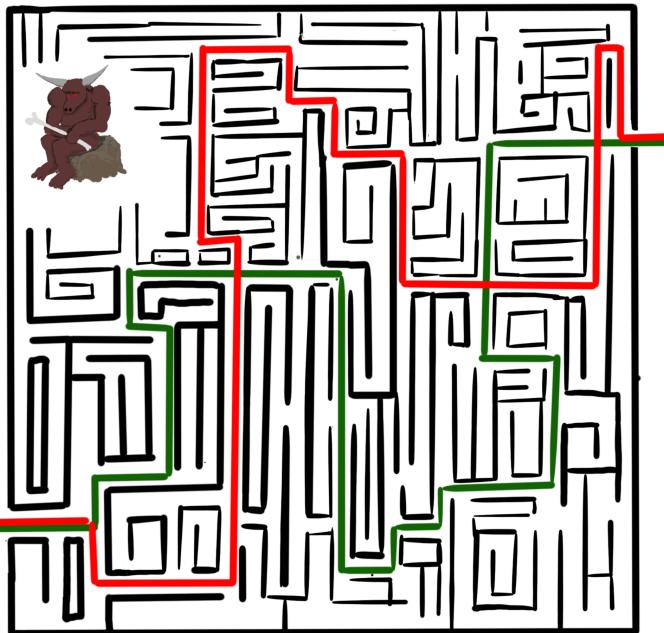
Of course, there are many technologies whose value was unknown *a priori* (personal computers, cars, etc.), so RDISPACs are not a silver bullet.

There are obviously a horde of unanswered questions. Some fun ones to think about might include:

- Is there a way to test public appetite for something like this?
- Publicly owned R&D assets have traditionally not done well — people are fine buying and holding when something is low volatility but R&D tends to be high volatility. Is there a way around this?
- What are the conditions on SPAC investments? There is still tons of business risk even if you hit technological milestones so this would be much higher risk than most things on the market.
- How can you both argue that a technology-based SPAC is a good investment and have its acquisition conditions be based only on technical milestones? Speculatively, the RDISPAC management could also do the legwork to line up advanced market commitments from governments or companies?

Again, this is wild speculation on a hype-filled trend. Like all the other speculations, it's hard to be sure that it will work but it's the sort of thing that's worth thinking about.

PARPA



Knowing things about the maze is worthless if you don't use that knowledge to chart a path. But as soon as you discover that what you thought would be a valid turn is in fact a dead end, the entire path needs to be rejiggered! As a result, this is weirdly both the most important and the most transient part of the piece!

In this part, I dig into the specifics of what *I* (and perhaps with your help, *we*) are setting out to do. If Part II was a description of the junctions that any maze-adventurer could face, Part III is the path that I expect my thread to trace.

I will first lay out the organizational and structures that I suspect will best dip, doge, duck, and dive between the constraints we explored in Part II. I'll then lay out a series of nested hypotheses that effectively make up the organization's roadmap. The notion is that 'testing' each hypothesis will progressively de-risk the giant spiky ball of assumptions underlying this whole scheme. These hypotheses will comprise three major phases in the organization's development. I'll outline these phases and the conditions for phase shifts.

Some of you have probably been asking "yes, but what programs will you

actually *work on??*" this whole time. Your appetite will finally be sated! In this part I'll sketch some potential programs that the organization could undertake.

Most people who enter a labyrinth — even those with a plan — don't survive. With that in mind, I'll close with a glance into the abyss and outline as explicitly as I can why this whole adventure could fail and what success could look like. My hope is that such an analysis will maximize both the chances of success and what we can learn in the unfortunate case of failure.

Precise visions are more likely to happen

A striking through-line that connects ambitious and successful technology programs from SpaceX to PARC (and ARPA computing programs before that) is that they were organized either explicitly or implicitly around a *precise* vision of the future.²³¹ These precise visions both act like a filter for likely-to-succeed programs and contribute directly to that success.

The causal relationship between precise visions and success is fairly intuitive. A precise vision acts as a coordinating mechanism that enables a large group of people to agree on what things they are working towards and, perhaps more importantly, what they are *not* working towards. Just as mythic rowers are useless without the coordination of a drumbeat, any compelling vision will generate action, but if it's imprecise the energy will quickly dissipate. It's the difference between rocket fuel creating a fireball and delivering a payload to orbit. If people can see the specific actions that they can do *now* in order to move towards the vision, it's comparatively easy for them to start doing them. However, most visions are abstract and fuzzy, which leads to reactions along the lines of "somebody should do something!" "Pass a law!" Or "it sure would be

²³¹ I want to acknowledge up-front that 'vision' and 'visionary' has become a grossly overloaded suitcase word. Unfortunately, it is also the correct word for what I'm talking about. I tried 'crisp picture' or 'shared goal' but they just don't work. It's also hard to define in a non-circular manner from the content of this note. Hopefully this excessive note serves to sufficiently discriminate precise visions from the casual use of the word!

nice if..." Legibility enables action!

I suspect that precise visions are also a filter for programs that are more likely to be successful. You could think of the vision like a test against reality. Do we understand the thing well enough to execute on it? If not, it's less likely to be feasible and may need more work before a program is likely to succeed. Of course, if you can precisely say what that 'prework' is, you've bootstrapped your way to a (different) precise program. Precision produces payoffs!

This litmus test doesn't limit you to just incremental changes either — Vannevar Bush's essay *As We May Think*²³² painted a precise vision of a thinking tool that was a massive paradigm shift away from anything that had existed before and yet precisely described many aspects of computers we'd recognize today.

Naming things is powerful

The men preferred to think they worked not in a laboratory but in what Kelly once called "an institute of creative technology." This description aimed to inform the world that the line between the art and science of what Bell scientists did wasn't always distinct. Moreover, while many of Kelly's colleagues might have been eccentrics, few were dreamers in the less flattering sense of the word. They were paid for their imaginative abilities. But they were also paid for working within a culture, and within an institution, where the very point of new ideas was to make them into new things.

— From *The Idea Factory*

Naming things is powerful. In some traditions knowing the 'true name' of someone or something gives you power over that thing. There's definitely something to that. Names are like nucleation sites for a concept-crystal to grow around; conversely, complex thinking about unnamed things tends to slide off of your brain. More pragmatically, naming things well is a useful tool for compressing and communicating concepts. It would just be annoying and weird

²³² See *As We May Think* (uses up one of three free articles at The Atlantic).

to constantly be referring to “this specific yet-unnamed DARPA-riff” even if we all knew what it referred to.

So what should this specific yet-unnamed DARPA riff be called? The working name was PARPA (for ‘Private ARPA’) Modern naming trends tend to

PARPA’s path through the maze is a series of nested hypotheses

A validated hypothesis is like getting to a junction you expected to be there, turning in the direction you expected to turn and not dying.

Bundled together, PARPA’s path through the maze looks like:

1. Create and stress-test unintuitive research programs in a systematic (and therefore repeatable) way.
2. Use that creditability to run a handful of research programs and produce results that wouldn’t happen otherwise.
3. Use that creditability to run more research programs and help them ‘graduate’ to effective next steps.
4. Make the entire cycle eventually-self-catalyzing by plowing windfalls into an endowment.

We can transform these steps into real action by making explicit the hypotheses each step entails. Each of these hypotheses is made of several constituent hypotheses. (And those are made of even smaller hypotheses until you get to things like “if I send an email to Professor so-and-so, we will have a meeting and they will be able to point me to two other people to talk to...” but I will leave that level of detail to your imagination.) I’ll present an overview below (with numbers and letters so we can refer to specific hypotheses) and then dig into each one.

1. It's possible to design programs in a systematic -and therefore repeatable) way.
 - a. It's possible to find people who want to do pieces of work that would not happen otherwise.
 - b. There exist several areas that could possibly yield results (defined loosely) in 3-5 years (given steady work).
 - c. For each potential program, it's possible to come up with a set of small experiments that could further confirm or deny its feasibility within 12-24 months and ~\$1M scale.
 - d. It's possible to get people to undertake seedling experiments.
 - e. Accomplishing 1a->1d will involve a combination of one-on-one conversations driven by reading papers, finding gaps, reaching out to people, and holding small, intensive workshops.
 - f. It's possible to find people besides myself who are excited to be PMs and actually *do* 1a-1e.
 - g. It's possible to systematize this process.
2. It's possible to execute on these programs in a way that relaxes academic/ startup constraints.
 - a. It will be possible to coordinate several research projects towards a coherent goal.
 - b. The early pieces of work will be done by some combination of academic labs, contract R&D orgs, independent researchers, and possibly people within corporate R&D who are able to do work for grants.
 - c. People will be willing to shift their organizational affiliation to an Exploratory Program Organization as the program goes on.

3. It's possible to 'graduate' these programs in a way that gives them a life of their own.
 - a. During the programs themselves it will be possible to do some of the pre-work to figure out the best way to graduate technology in order to maximize its beneficial impact on the world.
 - b. At least some of the people who have been working on the program will want to carry them forward in some form.
 - c. Some of the technology will make sense as a company, some of it will make sense as a non profit, some of it will make sense as a licensed technology, and some of it will make sense as open source.
4. It's possible for this entire cycle to eventually become self-catalyzing.
 - a. An early (non-monetary) component of self-catalyzation is a community that generates good inbound programs.
 - b. 'Pseudo-self-catalyzation' is state where the organization is getting enough consistent revenue through some combination of donations, spinoffs, and other sources like licensing and consulting that it can continue to run multiple programs assuming those revenue sources continue.
 - c. Full self-catalyzation can only happen by building up an endowment.

I'll walk through each of these in as much detail as possible. Fair warning: we'll quickly reach the limits of making research legible up-front. So many things need to be figured out on the fly and as a result the details of these hypotheses are going to read much more like working notes than a polished piece. Later hypotheses get more fuzzy because the farther you go in time, the more paths can fork. Many of the details about hypotheses sounds like a litany of ways that

it could go wrong — alas, this is the nature of these experimental things.

It's possible to systematically design programs

The first big question that any pragmatic person asks about PARPA is “so, what are you working on?”²³³ Building a process to consistently and precisely generate answers to that question is PARPA’s first major task.

“Figuring out what to work on” (or worse, figuring out how to figure out what to work on) as a first major step may sound a bit silly (and indeed it feels a bit silly to write it). Startups are generally created to go after a specific problem or opportunity so its a big red flag if they don’t know what they’re working on from day one.²³⁴ Conversely, actively “figuring out what to work on” is also a waste of time for most research funding organizations; once they’ve settled on a broad area, their modus operandi is to solicit and judge applications.

In both of these situations though, the ‘context’ of work is roughly fixed and well-established. People have patterns to match. A key reason for PARPA’s existence is to enable work that’s not being done because it doesn’t fall into those well-established contexts. Good PARPA programs will lurk in the shadows of counterfactuals, and counterfactuals are hard. Our job then is not just figuring out “what to work on” but three coupled questions “what things would be enabled by a new context, what is that context, and can PARPA create it?”

Program design will require answering all three in a way that resembles the final

²³³ “What are you working on?” Is an intriguing question. What counts as a legitimate answer to it is slippery. There are answers that are clearly legitimate like “we’re working on making webpage load times faster!” Unfortunately, “we’re working on a new institutional structure” is generally illegitimate. Perhaps it comes down to whether the questioner can imagine the sorts of day-to-day activities an answer would entail?

²³⁴ There are of course, many examples of startups drastically shifting what they’re working on. At any point in time though, good startups tend to know exactly what they’re working on.

step of solving a blacksmith's puzzle — each answer depends on all the others.

To give you a flavor of some of the questions we'll need to tackle in the process of figuring out what to work on: Is it a matter of coordinating multiple pieces of a system? Creating collaborations between disciplines? Prompting aggressive ideas that researchers aren't even thinking because they wouldn't normally be funded? Or does the thing not exist because it's actually just violating some laws of physics? What experiments could be done to test the answers to these questions?

Another reason figuring out what to work on is not straightforward: PARPA is especially susceptible to the chicken-and-egg situation facing any new institution: we can't be sure that we can work on something until we know who will be working on it, and people are hesitant to agree to work on something until the organization declares that it's working on it. We can't honestly declare that we're working on something until we at least have strong suspicions about who will be working on it. Unlike most startups that base their plans on hiring fairly fungible roles — "a software engineer" or "a data scientist"²³⁵, PARPA programs will hinge on working with people who have knowledge and skills that only exist in a few minds around the world. Many ambitious proposals start with "someone should do this crazy ambitious thing" but never actually specify who is both qualified and willing to do it. Finding those people and building trust-relationships with them will be a project in itself. Who is open to participating will shape the programs PARPA should undertake.

I've already begun this process for a few possible programs and frankly, it is a full-time-job in and of itself.

This chicken-and-egg situation will hopefully change once we have enough institutional momentum that we can count on inbounds, but even DARPA PMs tend to get precommitments from researchers to respond to calls for proposals before they launch programs.

²³⁵ This isn't to say all software engineers or data scientists are fungible! It is to say that the startup's ability to take even one step forward doesn't hinge on hiring the best software engineer in the world.



Finally, the entire process needs to be repeatable. Creating the initial programs will undoubtedly be ad-hoc and a bit of the mess, but in order for PARPA to be able to consistently create programs in the future it will be critical to spend time and effort paying attention to what works, what doesn't, and what could be systematized.

It's possible to find people who want to do pieces of work that would not happen otherwise

Most people don't advertise the fact that they want to go after slightly crazy research that requires resources or collaborators that aren't already at their disposal. This is doubly true if they actually have the skill to pull it off.²³⁶ There are exceptions of course, but most career-focused researchers file overly-ambitious ideas away 'for later.' Often these ideas are buried so deeply that they need prompting and trust to even talk about them out loud.²³⁷

A big part of the ICT's success depends on our ability to find these people. It will

²³⁶ There are plenty of people who have no idea what they're talking with great ideas that would change the world if only they could find someone to implement them.

²³⁷ The counterexamples that immediately popped into your head are the result of saliency bias. It's easy to recall people who openly want to work on things beyond their reach but obviously hard to recall the people who never talk about.

be a harder task than it might seem, especially without a track-record. While doing a few pieces of preliminary program design, I've found that the people who are excited about ambitious technical ideas often aren't in a position to execute on those ideas; either they don't have the technical expertise²³⁸ or because they're in a long-term position that precludes them doing the work. Conversely, the people who on paper are at the cutting edge of a field tend to have gotten there by being laser focused on a specific academic research agenda that can proceed methodically using the resources available to a single lab. This self-selected group is often distinctly unexcited by larger jumps that would involve combining their research with other disciplines.

Not all hope is lost! I have come across several people who *do* seem to have a magical blend of technical skills and excitement about ambitious possibilities that might be one step beyond what is consensus possible. These people tend to be post-docs, grad students, and technically trained people who have left academia. So the 'right' people do exist — the challenge is to figure out how to find them *consistently* and bridge the gap between 'intrigued' and 'on board.'

The people who would want to do early program work tend to be post-docs, grad students, and technically trained people who have left academia

After talking to dozens of researchers in the process of poking program possibilities, I've noticed a pattern. Tenured professors are happy to talk about the next step suggested by a specific paper and fuzzy long-term ideas, but post-docs, grad students, and technically trained people who have left academia ('tenuretrackless folks') are the ones who have that *precise* ambitious ideas that

²³⁸ I think credentialism is bad and that most people can retrain to do most things eventually if they try hard. However, success on a technically-hard non-software research project starting from scratch without guidance seems ... unlikely.

are critical for a solutions R&D program. It's frustrating to ask the point-of-contact for an exciting paper (who is usually the most senior coauthor ie. the professor) about what possibilities the paper opens up only to hear a list of more-of-the-same kind of experiments that are well within their grasp. If you can dig up the contact info for the tenuretrackless folks on the paper, on the other hand, you're often rewarded with "well, if we could simulate this better we'd be able to ..." or "if we could work with some physicists we could ..."

This response pattern makes sense: professors tend to become successful by finding a speciality and becoming its master through deliberate incremental work. This specialization makes them great at acquiring grants ("this person wants to do this feasible thing that nobody else in the world can do!") but pushes them into a mode where they have clear boundaries between the feasible within their realm of control and the infeasible outside of it. People with very little research experience tend to be the opposite: full of grand ideas but no idea how to implement them. Tenuretrackless folks are able to walk the middle way between the two extremes.

Another reason to seek out tenuretrackless folks is that they are (by definition) less tied to the traditional academic career path. So if a program ends up being successful, there will be a higher chance that they'll be open to 'graduating' with it to whatever form makes the most sense — whether it is a startup, nonprofit, or some kind of corporate absorption. Intellectual continuity is critical for a graduated program's eventual success in the real world.

Testing this hypothesis will entail deliberately skewing conversations and workshops towards tenuretrackless folks. This skew is tricky for two reasons. First, the person that people associate with an area of interest is usually a professor. Professor-obsession makes it hard to find the high quality tenuretrackless folks in the first place because one of the best tactics to find your way to interesting people is through recommendations. Second, tenuretrackless folks often don't have the same access to labs and equipment as professors, so working with them on externalized research will present extra challenges.

The focus on tenuretrackless folks should not be exclusive by any stretch of the imagination. There are tenured professors who want to use their tenure's freedom to its full extent and sometimes still do benchwork. There are high-school dropouts who learned to build bioreactors in their basements. However, I suspect those people will be such rarified outliers that the bulk of scarce resources will be better spent cultivating tenuretrackless folks.

The pieces of work in a PARPA program will tend to be things that people think can't get through a grant committee or require some sort of collaboration that isn't available to them

PARPA needs to nail down early the type of work that it's enabling. There are two opposing traps we could potentially fall into: on one side is supporting work that would happen regardless and on the other side is supporting work that is actually impossible. The delineation is clear as mud; there are even things that would happen regardless *and* are actually impossible. This nebulosity means that instead of rigid classification systems, we'll need to focus on developing heuristics.

I have a hunch that there are (at least) two heuristics for the type of work that falls into the sweet spot of 'possible but not incentivized': work that people suspect wouldn't be able to get through a grant committee and work would require some sort of weird collaboration. Both of these cases tend to fall into mental blind spots: instead of a nagging pain, they're the sort of thing where you think "that would be nice" but then forget about it. Part of PARPA's role is to tease out those discarded ideas.

Work might not be able to get through a grant committee for several reasons: it doesn't fit into a neat bucket; it isn't directly related to research that the

performer has done before²³⁹; or it just doesn't have strong evidence that it will succeed.

The sort of work that might require a non-existent collaboration is hard to generalize. An example might be a molecular-level process that a chemistry group wants to implement, but the time it takes to synthesize and the range of possible outcomes makes trial-and-error experiment infeasible. A machine-learning-driven simulation could narrow down the search space enough to make trial-and-error in the constrained search space feasible but the chemistry group has nobody with the expertise to build the simulation. At the same time, the simulation folks have nobody with the nuanced experience with chemistry to help make the simulations they're building actually useful. PARPA could fund both groups to work together and help coordinate between them.

Frontier-pushing technical people will be willing to work with a new organization

Even if it's possible to find people who want to do pieces of work that would not happen otherwise, there is a big gap between 'interested' and 'on board.' People will happily have conversations and spitball ideas; actually sitting down and doing work is an entirely different beast. This gap means that another hypothesis PARPA needs to validate early on is that we will be able to get researchers to work with us.

A big part of working with research is the ability to fund the research at all, which is what makes this hypotheses one of those experiments that can only be done in the context of an organization. However, money is necessary but nowhere near sufficient to get people to work with you. People need to be excited about the work; it will inevitably require both creativity and persistence in spades. We can't go all-in on "be excited! Pursue your passion!" though because any single project in the program because eventually the different

²³⁹ The tension between expertise and ambition strikes again!

projects need to converge. Finding this balance consistently will be hard!

Some tactics that we can test to ‘onboard’ collaborators include: small workshops of high-potential possible performers; spending a lot of work building trust with individual researchers;²⁴⁰ working to build a long-term community that will continue to be valuable to researchers after any specific project.

There exist several areas that could possibly yield results (defined loosely) in 3-5 years (given steady work)

PARPA’s success depends on creating several programs in the ‘sweet spot’ where they’ll yield ‘tangible’ results in 3-5 years. ‘Tangible’ here doesn’t necessarily mean a product per se, but an artifact that inspires confidence in people outside of the organization. This constraint is purely pragmatic — realistically, even people who are excited about PARPA’s potential (hopefully you, dear reader) will lose enthusiasm if the organization doesn’t make some sort of visible progress over several years. As the organization builds more trust over time, it will be able to go after longer-term programs and work on promising projects for “as long as it takes,” but the path-dependence is important.

This plan rests on the hypothesis that the sweet spot even exists. It could be the case that solutions R&D programs that could show results in 3-5 years are uninspiring and/or are already being tackled for any number of reasons — a frothy VC market, unusually prescient grant allocators, actual DARPA, etc. This is where program design plays an important role: it’s important to not only have reasons to believe that a program will be successful, but to early on have reasons to believe that there’s a chance that success will happen on a useful timescale. The two specific programs that we have been digging at so far are General-

²⁴⁰ Warren Weaver, the Rockefeller Foundation program officer who funded early genetics work expands on this point in his notes for other program officers.

purpose telerobotics and first steps on what would be a series of programs towards atomically precise manufacturing. We're open to suggestions!

For each potential PARPA program, it's possible to come up with a set of small experiments that could further confirm or deny its feasibility within 12-24 months and ~\$1M scale

DARPA PMs use seedling projects to 'acid test' the riskiest pieces of a program idea.²⁴¹ I'm confident that PARPA will be able to do something similar for our programs but it's still a hypothesis we need to explicitly test. It is extremely hard to call out in advance which pieces of evidence would compellingly suggest that a crazy idea is at least possible.

Part of derisking this hypothesis is frankly just a learning process about what makes a good seedling project. After studying examples of past DARPA seedling programs, some potential heuristics are:

- Good seedling projects specifically go after the reasons that everybody says that something won't work. Their goal isn't to make it work, but to at least show that the reason why it won't work isn't the game-ender that it at first appears to be. For this approach to work, it's critical to precisely pin down both *what* won't work and *why* experts assert that's the case.
- The tagline will either sound incredibly silly ("How memes become reality") or incomprehensible ("Augmented Reality-based Collaborative and Corroborative Collection"). Speculatively, this heuristic might work because seedlings need to either be poking at a broad hunch (that it might be possible for memes to become reality) or a specific interesting technical thing that you suspect could become significant.

²⁴¹ See the section DARPA PMs use seedling projects to 'acid test' the riskiest pieces of a program idea from Why Does DARPA Work?

I suspect the line between success and failure in seedling projects will often be fuzzy. This nebulosity is why developing trust with performers will be critical. Human judgement will be unavoidable in deciding whether the seedling experiments provide enough evidence to launch a full program. Another possible frame for the seedling programs is that they serve as a trust-building ‘trial’ period. Think of internships or how some companies set it up so that employees and the company can both choose to part ways after a few months.

Below are some top-of-mind examples²⁴² of speculative seedling projects to prime your intuition pump. Binding constraints They might not be feasible for many reasons!

- Showing that it might be possible to get muons to have a higher replication rate in a fusion reaction.
- Doing a study on what that design for a broadly-applicable robot specifically designed for telerobotics work might look like.
- Experimenting with the binding constraints on scaling up carbon-free cement.
- Testing AlphaFold in a closed-loop with specified properties for the blocks that a 3D-protein printer would use.
- Showing that you can make a nanoscale breadboard by applying implosion fabrication to a DNA microarray.
- Showing that temperature-hardened electronics and communications developed for space missions could withstand the conditions involved in drilling 20 km into the Earth for geothermal energy.
- Designing a MEMS device that could polish/sharpen its own features after fabrication.
- Prototyping a small sensor that can extract DNA from muddy water while

²⁴² The last four are directly from Adam Marblestone.

harvesting the mud as a biofuel and allowing self-sterilization / cleaning.

It's possible to get people to undertake seedling experiments

An obvious but important challenge is to bridge the gap between potential performers being ‘interested’ and actually taking action. How do we shift the people we found while testing the hypothesis that “it’s possible to find people who want to do things that would not happen otherwise” from interest to action? It’s one thing to find people who have both the ambition and skills to undertake things and another to actually get them to spend the time undertaking them!

This is where having money and an organization is actually important. At some point we need to be able to say “we’ll fund the work to explore this thing.” Does there need to be a contract? What sort of agreement does it look like? Can we work directly with the tenuretrackless folks or will we be forced to go through a professor or university bureaucracy? Could we ‘hire’ someone for an ‘internship’ to do seedling projects? There are fractal questions that will only be resolved by just doing the thing.

Accomplishing hypotheses 1a->1d will require a combination of gumshoeing and workshops

Finding people and promising areas, figuring out the experiments they could do to poke at binding constraints, and then pulling everything together so those experiments happen will require a lot of fuzzy ‘people-wrangling.’ If you write down what the work I would call ‘gumshoeing’ entails — lots of emails, talking to people, following up, connecting, circling back, negotiating — it certainly deserves the uncomfortable side-eye it gets from people (like myself) who like concrete actionable plans. Doing those things can easily lead nowhere if done poorly. Hopefully, by explicitly calling out the possibility we can avoid that trap,

but it remains a deeply unsatisfying answer.

Workshops are a systemized way of people-wrangling

DARPA program managers often hold small workshops of experts as part of their program design process.²⁴³ Running a series of these workshops might be a small, actionable place to start PARPA.

I suspect that big part of starting and testing PARPA's viability is to design sufficiently precise programs: External researchers need to believe the projects in the program are a convincing use of their time; external funders need to be willing to provide the money to start executing on those programs; and we need to believe that they actually have a chance of unlocking something amazing if they're successful.²⁴⁴

One way to start designing programs is to run a series of workshops where a handful people at the edge of a discipline with different backgrounds get together to riff on each other, explore, and argue their way towards a clear picture of the edge of possibility.

What would a workshop look like concretely?

Imagine 5-7 people who understand the edge of a discipline but are fringy in one way or another getting together in person in a big whiteboard-filled house in the woods somewhere with relatively easy access to an airport. The participants would spend a week digging into the precise problems in an area like "why don't we have black box factories?", "how could we grow infrastructure like plants?", "how far could we push our ability to do positionally-controlled chemistry?"

²⁴³ See the section A large part of a DARPA program manager's job is focused network building from Why Does DARPA Work?

²⁴⁴ See question #4 of The Heilmeier Catechism: "Who cares? If you are successful, what difference will it make?"

with the goal of establishing what's possible in a ~2-5 year time frame that's not being worked on, precisely why not, and what work would need to happen in order to get there.

Admittedly, it's a big ask for people to take a week out of their life. Unfortunately, less than a week is probably insufficient. It takes people a few days to get comfortable with each other to the point where they're willing to dig into out-there ideas. It *might* be possible to boost the comfort level beforehand with digital tools, but I'm skeptical.

Workshops would ideally produce several concrete outputs

- A list of preliminary seedling projects that could act as go/no-go experiments for a larger program along with who should be doing them.
- Answers to all of the questions in The Heilmeier Catechism or precise reasons why the answer is unknown or nebulous (and the project you would need to do to firm it up).
- A 'roadmap' of what would happen in a full program if the seedling projects pan out.
- A list of key people in an area to loop in on the project.

Optimistically, after running a few workshops, the participants could form the core of a larger community. This group would ideally serve two purposes: it would help people who want to push the edge of their fields find overlapping and complementary people²⁴⁵ and be a pool of people in PARPA's orbit who could help generate program ideas, act as contractors, and perhaps become PMs or performers. As a proof point, Ink&Switch seems to have pursued this strategy successfully!

Pros and cons of hub-and-spoke model of program design

²⁴⁵ In Reinventing Discovery, Michael Nielsen elegantly describes this process as 'designed serendipity'

The alternative to workshops would be for the Program Manager to talk to a slew of experts directly and do the synthesis themselves. This the ‘hub-and-spoke’ model of program design has the PM at the middle doing the majority of the talking with the experts and perhaps connecting them as needed. This approach has pros and cons. The pros are that is that a PM will be maximally bought in to a program entirely of their own creation and the idea will be the maximally coherent. Committees tend to lead to median results and it’s easy for workshops to devolve into committee-thinking.

On the other hand, the hub-and-spoke model requires the PM figure out all the right questions to ask or play the go-between for different experts. The hub-and-spoke method might be ideal when a PM comes in with a strong hypothesis about the program and expertise in the area. I might be able to do this for a simulation or robotics-based program but would be unable to do it for something in biology or materials.

It seems important for PARPA’s initial programs to be relatively uncorrelated. Programs designed via the hub-and-spoke method by the same PM will be more strongly correlated with each other than via the workshop method. Early on, we won’t have very many PMs, which could exacerbate the correlation. These facts suggest that while it’s worthwhile to pursue both strategies, the potential pros of the workshop method and the potential downsides of the hub-and-spoke method mean that PARPA should lean more heavily on workshops early in its life.

It's possible to find people who are both willing and capable to be Program Managers

Like any DARPA-riff, PARPA’s success or failure will hinge on our ability to work with amazing program managers. It’s not a given that we’ll be able to find the right people and persuade them to work with us. Pickiness will abound on both sides: people who would be good PMs have many options in life; at the

same time, PARPA needs to be discerning about who becomes PMs because of the trust and responsibility that they must shoulder.

Given these challenges, how do we propose to find PMs and why will they join?

A chunk of this hypothesis is based on confidence that the frustratingly vague ‘people-wrangling’ that also goes into program design will yield fruit. “I suspect that if I try really hard I will succeed.” This hypothesis is not entirely based on hubris though — below are some more precise hunches about where to find PMs and what will convince them to forgo other options:

- The same ‘tenuretrackless folks’ who might make good performers may also make good PMs. Performers often become PMs at actual DARPA.
- People who are thinking about starting companies (or started and stopped) as the best of bad options for enabling a technology would be good PMs. Some subset of people who start startups are doing it not because they want to build a business but because it seems like the best way to get a technology into the world. In some of those cases, a PARPA program would be far more effective than a startup.
- While we won’t be able to pay as much as a big tech company, I suspect that the people who would make good PMs aren’t out to maximize their income but instead have some income threshold, above which the non-monetary benefits of the work (which is uniquely excellent!) dominate their mental calculus. Without being tied to any government pay scales, we can negotiate this threshold with individuals.

It's Possible to systematize the whole program-creating process

Validating this hypothesis looks like getting to a place where after we’ve created some hopefully-not-too-large number of programs, we (internally at least) feel confidence that it wasn’t just a fluke. Ideally, we would also be able to

instill that confidence into people outside the organization as well. The ability to repeatedly create good programs is a strong reason to believe that it might be successful in the long run.

A good chunk of systemizing program-creation will entail nothing but explicitly paying attention to what works and what doesn't while going after the previous hypotheses. However, I also suspect that more abstract work on program design as a discipline will also be important. Perhaps you could look at them as the empirical and theoretical sides of the same coin.

It's possible to execute on programs in a way that relaxes academic/startup constraints.

Program designs are useless if they don't translate into useful outcomes. People won't just see a roadmap and leap into action — there will almost by definition be significant friction since we're explicitly going after programs that we suspect wouldn't happen otherwise in the current system. As a result, getting programs done will be a challenge!

It's easy to imagine scenarios where we run workshops, create precise ideas about the riskiest questions in them, know the exact right people to work on different parts of the program, and even do some seedling experiments to show that what we're proposing is not impossible and yet fall flat on our faces when it comes to a full program. Coordinating multiple long-term research projects and synthesizing them into a working technology is an entirely different beast.

A clear failure mode is that we fail to achieve escape velocity from existing incentive structures. Maybe all the people who have the experience to do the work still want to optimize for publishing papers. Maybe the program requires coordination between a number of disparate groups and all of those groups want to optimize their particular piece instead of accepting constraints that make the entire program eventually successful. Maybe there are legal problems with the researcher's organizations. Maybe the people working on the program graduate

or get a job offer halfway through. The list goes on.

The overarching hypothesis is that we'll be able to avoid these known pitfalls, discover and avoid the unknown unknowns and successfully execute on full programs. Derisking program execution entails several sub-hypotheses:

It will be possible to coordinate several research projects towards a coherent goal

Coordination is hard. Research is hard. Coordinating research is going to be extra hard. Dealing with dependencies between projects will be a challenge both on the technical front and the human front. Technical coordination will require figuring out how to keep each project going despite hiccups in other projects. Human coordination will require making sure people communicate and get along. These are normal management problems but they will be turned up to 11. The distributed nature of early programs will compound with the fact that doing anything new (especially with atoms) makes it harder to have shared context. At the same time, shared context will be critical!

To illustrate what the coordination problems might look like, let's look at an imaginary program to use molecular machines in metal-organic frameworks to do one shot molecular synthesis. There might be one project to create better simulations for a molecular process, one project to design a series of molecular machines that could potentially turn a multi-step synthesis process into a single-shot process, and one project to synthesize the system. Friction can pop up all over the place: benchmarks for the simulation need to be processes that are actually relevant for the second group; the design group and the implementation group need to avoid both designing something that is actually impossible to create and also avoid limiting themselves to conservative designs that would be straightforward to create.

There are many ways to potentially address these coordination problems, all of which need experimental de-fuzzification. PMs can check in regularly with each

group, not to crack the whip about progress but to check on what each group needs from the others and whether they're getting it. One of the PM's roles will be keeping the bigger program goal front and center and figuring out when an individual group needs to do something 'suboptimal' in order to maximize the entire system. Another tactic is to regularly bring together the people actually doing the work on each project (as opposed to just the PIs) *before* they have results so that people across the program can build rapport talking about real research problems instead of defending results.

The early pieces of work will be done by some combination of academic labs, contract R&D orgs, independent researchers, and possibly people within corporate R&D who are able to do work for grants

This hypothesis has three implicit sub-hypotheses: 1) the early work in programs will be done in distributed external organizations 2) the types of groups that will do that work will be heterogeneous — that is it won't just be academic labs 3) PMs should be agnostic to the types of organizations doing the work as long as they can get it done.

Any one of these sub-hypotheses could be wrong! Other private organizations like Ink&Switch and various open source projects have pulled off distributed, externalized research but that may not translate to atom-based solutions R&D. It may turn out that externalized non-computer-science research only works when you're a government organization with massive resources that can buy work at contract research organizations like SRI or APL. Academic labs might have draconian licensing or university overhead requirements that rule them out as potential partners. Independent researchers, regardless of their ability, might not have access to the equipment they would need to undertake a project.

However, I'm optimistic that there will be ways around these potential problems!

People will be willing to shift their organizational affiliation to an Exploratory Program Organization as the program goes on

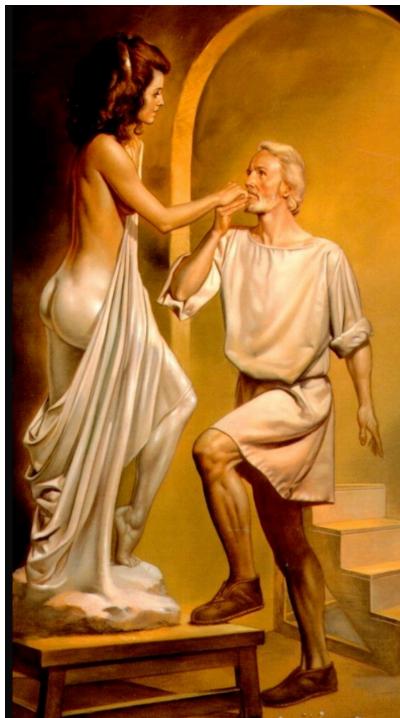
This hypothesis exists at two levels: "would people from different projects in a program be willing to shift into a single organization?" and "is consolidating into a single organization even a good idea?" I lay out the reasoning for why Exploratory Program Organizations might be a good idea in Part II so I won't dig into it here. To the question of whether people would be willing to join the EPO, like most human systems there are many factors at play and it's impossible to predict all of them a priori. How much people come to care about the program they're working on would ideally be a deciding factor but in reality it's a complicated mix of incentives: what career track they see themselves on; what the opportunity cost of joining is; how they weight chance the chance of a good outcome; etc. Managing all of these incentives will be a major challenge.

It's possible to 'graduate' programs in a way that gives them a life of their own

"Manage the transfer, not the technology: Innovative leaders with some successes tend to appoint themselves loonshot judge and jury (the Moses Trap). Instead, create a natural process for projects to transfer from the loonshot nursery to the field, and for valuable feedback and market intelligence to cycle back from the field to the nursery. Help manage the timing of the transfer: not too early (fragile loonshots will be permanently crushed), not too late (making adjustments will be difficult). Intervene only as needed, with a gentle hand. In other words, be a gardener, not a Moses."

— Safi Bahcall, Loonshots

Even if PARPA manages to nurture programs all the way from baby seedlings to impressive results, all that work will be worthless if the programs don't have a (positive) impact on the world. In order for that impact to happen, the programs need to take on a life of their own. There are many ways that technology can take on a life of its own but unfortunately none of them are as easy as just kissing it and hoping the gods do the rest. It could end up as an 'open' project, be carried forward by a nonprofit, absorbed into an existing organization or become the core of a startup. Regardless of the outcome, I suspect that explicitly thinking about which outcomes to target and course-correcting towards them will maximize the viability of programs after being part of PARPA.



Bringing technology to life is harder and hopefully less awkward than this.

One sub-hypothesis of 'graduating programs' is that there is a lot of 'pre-work' that can be done during the course of a program to maximize its impact after it concludes. It would be easy to focus monomaniacally on getting to some

tremendous milestone, reach it, and then look around and ask “ok, what do we do with this now?” Or worse, assert “well, our job here is done, it’s up to others to carry the torch forward!” Instead, PMs could act a bit like curling sweepers — altering the friction in front of the program ever so slightly so it ends up in the best place possible.



This ‘pre-work’ will involve going back and forth with potential manufacturers (“what would make this technology actually producible at scale?”); potential adopters (“what specs do you actually care about? Why don’t you use something like this already?”); and regulators so that the previous two groups aren’t worried. The PM/program team will obviously need to make a call on what to incorporate into their work — throw out “I want a faster horse!” But keep the “power/weight ratio doesn’t matter to us but power/volume does.” There will also be hard questions around what ‘form factor’ would maximize diffusion — is it a product? A process upgrade? A lab that does a particular kind of microscopy or synthesis work? Does it perhaps need to be the starting point for another program? A lot of this work closely resembles standard startup best practices — I

imagine them both as ‘finding a fit in the real world’ but where a startup is trying to optimize long-term capturable value through a scalable product a PARPA program has more degrees of freedom.

Focusing too intently on what will happen after the program ends could have deleterious warping effect as well! As we’ve beaten to death already, one of the reasons PARPA needs to exist in the first place is because existing institutions focus work on outcomes like papers or profit. Navigating between the drive to enable potentially impactful work and nudging that work so that it actually delivers on that impact is yet another unavoidable tension that PARPA must confront.

Another sub-hypothesis is that in order to successfully ‘graduate’ a program needs some of the people who were working on it to carry it forward. Intellectual continuity matters. And of course, we suspect that some of the program participants will *want* to graduate with the program. This intellectual continuity could take many forms — consulting with other organizations implementing the technology, starting a startup or nonprofit, perhaps something less conventional. The key thing is that at least one person who worked on the technology full-time *continues to do so*. It’s a common misconception that ‘technology transfer’ can happen by handing off some patents and perhaps some light part-time consulting work. This might have been the case in the past when patents were schematics of a mechanical part but no longer! As a result, it will be important for at least a few of the people in a (temporary) PARPA program to see it as long-term endeavor.

PARPA will eventually spin out companies

Spinning out²⁴⁶ as a company is only one of several ways that programs can

²⁴⁶ I prefer the imagery invoked by ‘budding off’ companies instead of spinning out, but the convention is strong with this one.

'graduate.' However it's worth talking about them directly because startups dominate so much of the cultural zeitgeist and potentially play a unique role in PARPA's business plans.

At the end of the day, PARPA's goal is roughly to unlock more technologically enabled wonder in the world. The "In the world" part does a lot of work. It demands technological diffusion, which can happen in one of three (not-mutually-exclusive) ways: you can sell technology directly; you can give it away; or you can build an organization around it.

Some of PARPA's work will certainly be appropriate for selling directly to companies that might diffuse it. However, there will be many things that just don't fit into the existing product lines or business models of current organizations.

Giving away non-software technologies tends to have limited diffusive potential outside of an institutional framework. Most people don't make their own stuff, so if you just put designs on the internet, they're unlikely to be widely used. There are of course counterexamples, but more often the diffusion of free or open technology is eventually driven by organizations. Penicillin needed Pfizer; the semiconductor needed Fairchild; electricity needed Edison Electric.

The importance of companies for diffusing technology means that PARPA will (sometimes) need to spin out companies *in order for it to achieve its goals*.

There are, of course, caveats: Not all of these companies will be high-growth startups. Even though a technology might be at a point where it needs a company to carry it forward, that doesn't necessarily mean that company will be a good venture capital investment. The appropriate structure for some of these companies might even be as nonprofits, like the Mozilla or Wikimedia foundations.

It's possible for this entire cycle to eventually become autocatalytic

The longest-term, vaguest, and most aggressive hypothesis is that it's possible for this entire cycle of program design, execution, graduation to become 'autocatalytic'; that is, PARPA can (and must) become its own money factory. Autocatalyzation isn't just about money, but something bigger: a cycle where good results beget good results.

Why does autocatalyzation matter? The underlying meta-hypothesis is that PARPA will only be able to achieve truly miraculous things if it plays a long game. Good research has long timescales — it took the industrial labs of the 20th century decades to hit their stride. It's fairly unsubstantiated but it feels like an organization's timescales need to be even longer than the work it undertakes. There are also compounding returns to institutional competence in high-variance domains. Of course, this meta-hypothesis could be wrong as well! There are several research organizations that were flashes-in-the pan but produced outsized results: The Rad Lab and Willow Garage leap to mind.

I suspect that in order to thrive in the long run, PARPA needs to be as in-control of its own fate as possible. If PARPA never stops depending on external capital, it will eventually fall into one incentive trap or another. Even with the best funders, the wisest leaders, and the most robust organizational structures, if you roll the die enough times you'll eventually roll a natural 1. Of course, autocatalyzation does not guarantee success, but it seems like an eventually-important part of it.

We can divide hypothesis that *PARPA's 'core gameplay loop' can become autocatalyzing* into three sequential sub-hypotheses:

1. PAPRA can create a community that generates good inbounds and enables more successful programs. The first hint that PARPA can become self-sustaining may not be monetary at all. In the early days, we will need to spend significant effort hunting down good performers and programs. If PARPA is doing good, differentiated work this should shift over time and we should be

able to become a honeypot for excellent people who want to work with us and ideas that couldn't live anywhere else. You can tell a lot about both people and organizations by the company they keep. An autocatalyzing community is of course insufficient for an autocatalyzing organization. However, it may be necessary and will happen more quickly than monetary autocatalyzation, thereby providing an early 'test' of the autocatalyzation hypothesis.

2. PARPA can get to a 'Pseudo-autocatalytic' state where the organization is getting enough consistent free cash flow through some combination of donations, spinoffs, and other sources that it feels 'default alive.' For long while PARPA will be living pretty hand-to-mouth — we'll be able to raise money here and there but they will be one-off events. At some point after demonstrating promising results (and probably after creating an autocatalyzing community) funding will become dependable and / or large enough to create a small buffer. People will donate regularly or larger amounts; spin-offs might have a few successful exits, encouraging more people to invest; we could establish a consortium or consulting agreements that bring in revenue. At this point we will have flipped from 'default dead' to 'default alive'.²⁴⁷ This state will still be full of incentive traps and therefore long-term unstable but will nevertheless be a significant step.

3. Full autocatalyzation can only happen by building up an endowment, which will be possible by squirreling away excess from the pseudo-self-catalytic process. An endowment that funds operations out of the interest on a large principle seems like the ideal funding mechanism for a private DARPA — the question is whether building one is a chimera. Our working hypothesis is that building an endowment will be extremely hard and likely to fail but isn't impossible. It will take a long time and strong financial discipline; as I noted in the section on endowments, it's a bad idea to try to build it all at once so it will need to be built up slowly over time by limiting expenses to a portion of the organization's revenue. A big assumption here is that funders and investors will

²⁴⁷ See Default Alive or Default Dead

be on board with this plan!

PARPA can and should create a replicable institutional model

There are several attributes of solutions R&D and PARPA that suggest that in order to maximize awesome in the world, we should explicitly think about how to create a replicable institutional model while building PARPA. The model is inherently unscalable and institutional models are generally more robust than any individual institution. This final hypothesis is more of a meta-hypothesis that we should be paying attention to throughout every stage of the organization.

While PARPA will hopefully be able to accomplish many important things, it will be structurally unable to do *all* the important things. Companies can conceivably²⁴⁸ monopolize an entire market and provide all the search, steel, cat videos, or diapers in the world. By contrast, DARPA's tiny and flat structure likely plays a key role in their success²⁴⁹ and any organization riffing on its model will see diminishing returns beyond its size.²⁵⁰ This may be similar to what you see in venture capital firms and many private equity firms — small firms are able to get crazy returns on outlier results while large ones revert to the mean.

Blazing the trail for new institutional structures is far more impactful than any given institution. The impressive thing about Sequoia and Don Draper is less the investments they made, but the fact that they created a template for an entire institutional model.

Contingent factors can always kill a given organization even if it gets the model

²⁴⁸ Though they rarely do in practice.

²⁴⁹ See the section DARPA is relatively tiny and flat Why Does DARPA Work?

²⁵⁰ 1974 Bell Labs and 2020 DARPA have uncannily similar budgets (\$3.6B and \$2.8B in 2020 dollars) - there is not clear causal linkage but it is suggestive.

correct. Or perhaps like ARDC²⁵¹, arguably the first VC firm, you get the idea and some pieces correct but screw up others (like legal structures) in an ultimately fatal way.

What would you differently if you're trying to create a replicable model vs. just building a one-off? One piece is to consciously document decisions and failures — why you do one thing over another and things you tried that didn't work. Additionally, it means forgoing a level of secrecy and being open to helping people who want to do something similar.

PARPAs ideal legal structure comprises both a Nonprofit and a for-profit

Note: Many of the decisions in this section are based on numerical assumptions. I've marked each of these assumptions with a . I've put these quantitative factors into a model that can be referenced [here](#), so you can copy it and play with it yourself.

Another Note: This is a description of a hypothetical platonic ideal — there are surely things in this section that are either illegal or pragmatically impractical because key people will say "ha no I'm not signing that." However, it's

The ideal legal structure for the ICT would satisfy three conditions:

1. The legal structure should enable people to make tax-deductible donations to the organization but also be able to channel for-profit investments towards valuable work
2. The legal structure should enable the organization and the people involved in it to participate in *some* upside of capturing value from technologies they've created.
3. The legal structure should enable the organization to become self sustaining. A self-sustaining organization either needs to sell a product or

²⁵¹ For much more about ARDC see Creative Capital: Georges Doriot and the Birth of Venture Capital

service, have an endowment, or regularly generate liquidatable equity.

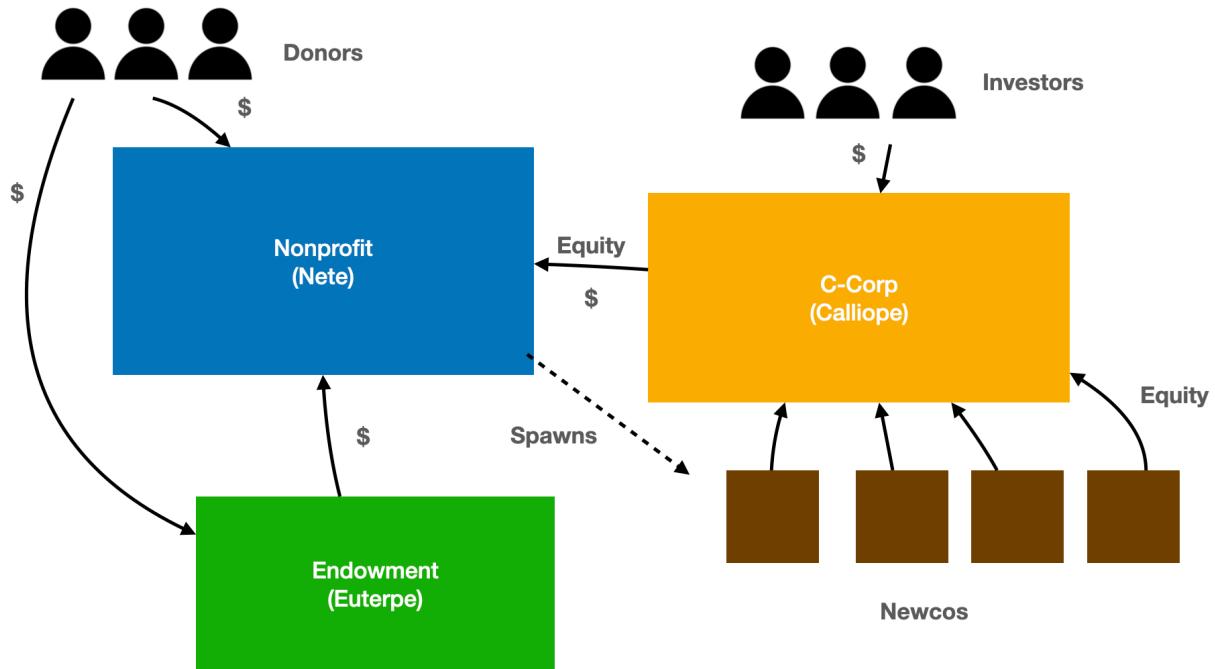
On its face, these three conditions seem contradictory. In aggregate we can't expect an organization that seriously tackles solutions R&D to be profitable. Especially if it focuses on the work that the relatively efficient market in for-profit organizations won't touch.

However, while an aggregate of all of PARPA's activities will not be profitable, it's a reasonable (but still risky) hypothesis that a portfolio of companies spun out of PARPA could end up doing well. Regardless of profit, PARPA will eventually spin out companies because companies can be an effective mechanism for technology diffusion. It's an even safer bet that if PARPA is doing something useful, there will end up being many profit-generating organizations that can trace their DNA back to PARPA. But let's ignore the latter for now.²⁵²

This will be a sketch of the simplest structure that (I believe) meets all three criteria — let's call it the platonic structure because implementing it will undoubtedly involve some changes for pragmatic reasons like "not going to prison on a technicality." In a sentence, you have a big N-Nonprofit that owns a c-corp and also has an endowment. For convenience let's name them after the muses Nete, Calliope, and Euterpe because it will make the section much more interesting than saying "The Nonprofit" "The c-corp" and "The Endowment" over and over again. Nete runs all of PARPA's activities. Her operating capital comes from three places: tax-exempt donations from donors, cash from Calliope, and dividends Euterpe. Calliope owns parts of companies that spin out of Nete either as equity in the case of high-growth companies or as profit-share agreements in the case of low-growth companies. People can invest in Calliope in exchange for a claim on future profits. Finally, Euterpe does what endowments do — she invests money broadly and attempts to generate a steady dividend that she gives to Nete. Euterpe gets money either directly from donors or from the

²⁵² If every company whose core business depends on technology that can be traced back to DARPA gave some tiny percent of their revenue to DARPA, it would fund DARPA many times over. (And in a way they do, via taxes in the US.)

Nete when her revenues exceed her expenses.



At a high level this structure satisfies criteria #1 by enabling people who prefer to make tax-deductible donations to donate either directly to Nete's operational budget. The structure satisfies criteria #2 by enabling people to invest in only the outcomes of programs that eventually become companies, which I do suspect could create real return, as opposed to the structure as a whole. Finally, the structure satisfies criteria #3 through the fact that Nete is the "parent org" and *a reasonable hypothesis about how Euterpe can get to a point where she can sustain operations.*

There are a number of nuances, especially around Calliope, the c-corp. The first nuance is about how Calliope's stake in spin-out companies works. We've already established that PARPA will eventually spin out companies. However, it's important to reiterate both that not every successful program will turn into a company and that not every company will be a high-growth venture-backable startup. Sometimes the best vehicle to get a technology out into the world is a small, specialized organization. Standard equity ownership does not make sense for slow-growth companies because they may never be acquired or go public. Instead, Calliope could own a share of their future profits. This profit-sharing

will end up being negligible on the for-profit's balance sheet in a best-case-scenario  but it is a useful source of free cash flow. So in order to maximize technology getting out into the world, it's important to offer companies a choice of profit- or equity-based ownership. The size of that ownership is another tricky matter.  The more ownership Calliope maintains in the baby companies, the less profit or equity they'll have to fund their own operations and the less likely they are to succeed. At the same time, Calliope needs to own enough to make a return for her investors. As anchors, Y Combinator takes 7% and Entrepreneur First takes 10%. *We'll go with 10%* because it's a nice round number and any higher number feels like it could cause significant damage to a startup. Perhaps it should be lower.

The next nuance is about Calliope's legal status. The two main choices are whether she should be an LLC or a C-corp. The most pertinent differences between the two revolve around taxes and the ability to issue shares. LLCs are 'pass-through' legal structures so only investors would be taxed on income they make from Calliope. As a C-corp both Calliope *and* her investors would be taxed. C-corps have much more legal precedence and people are more comfortable with them, so unless the double-taxation is a significant burden it's generally a good idea to default to a C-corp. The way to navigate this tradeoff is to look at where Calliope's (hypothetical) value is going to come from and which structure would maximize technological diffusion (do not forget our goal!) The majority of Calliope's value will be in holding equity in companies that are high value. The classic trap here is to set up Calliope's incentives so that she is incentivized to lean on those companies to have a liquidity event as soon as possible. Calliope would be incentivized to push for liquidity events if her value were in the cash she would be returning to investors. Instead, it would be better if Calliope's value was based on the long-term value of the equity she held. Aligning those incentives seems to suggest that Calliope should be a C-corp. While this might be an obvious conclusion to you, it was not actually my first assumption. The one other consideration that might tip the balance is that there are more constraints on how ownership works in an LLC than a C-corp.

The next nuance is how ownership of Calliope will work. The ways that you can legally divide up ownership in an organization are heavily regulated so I am going to first describe a platonic system that is probably illegal and then explain how it could be crudely implemented within the constraints of the law.

Obviously investors and Nete herself are going to own a good chunk of Calliope, but it's also important to figure out how to enable Nete's PMs to participate as well. Enabling PMs to own a meaningful part of Calliope is actually important to incentives, when you consider that there will be some programs that will naturally lead to valuable companies but many others that will be hugely impactful, but won't make good companies. If the implicit choice for PMs is "if your program becomes a company, you'll get financial upside but if it doesn't you get nothing" you will incentivize PMs to either push programs that shouldn't become companies towards becoming companies or to never start uncommercializable programs in the first place. If instead you give everybody some slice of the value Calliope captures, it will hopefully alleviate that pressure. Of course, this won't stop PMs who are strongly driven by economic incentives from skewing all their work towards companies. Nor will it create totally equal outcomes — PMs who go on to create successful companies will obviously own far more of those companies; arguably the majority of a technology company's "excess value" is not actually due to its technology.²⁵³

Ownership based on an accumulating 'ledger' of points would enable everybody involved in the organization — both those who put in money and those who put in time — to participate in upside on roughly equal ground. You could assign different number of points for different activities that help the enterprise as a whole: "you get N points for the first year of service, 1.2*N for the next year, etc." "You get M points / \$100k." At the end of the day, you end up with a list of people and how many points each of them has — ownership is just your points divided by the total number of points. The ledger system ultimately has the same outcome as selling and awarding people shares that get progressively diluted as a company increases in value (more points will be created over time so a fixed

²⁵³ See Productive Uncertainty, again.

number of points will represent a decreasing fraction of Calliope). However, a point-system has several positive features that standard shares do not and avoids some of their downsides. One upside to this system is that it gives everybody the same type of ownership - avoiding preferred shares, etc. The ledger can roughly stand as an accounting of who contributed to the organization's success.²⁵⁴ Combined with 'pseudopoints' for donations to the nonprofit (which can't be financially rewarded) the ledger could be a public source of pride instead of a secret and source of jealousy like company ownership normally is.

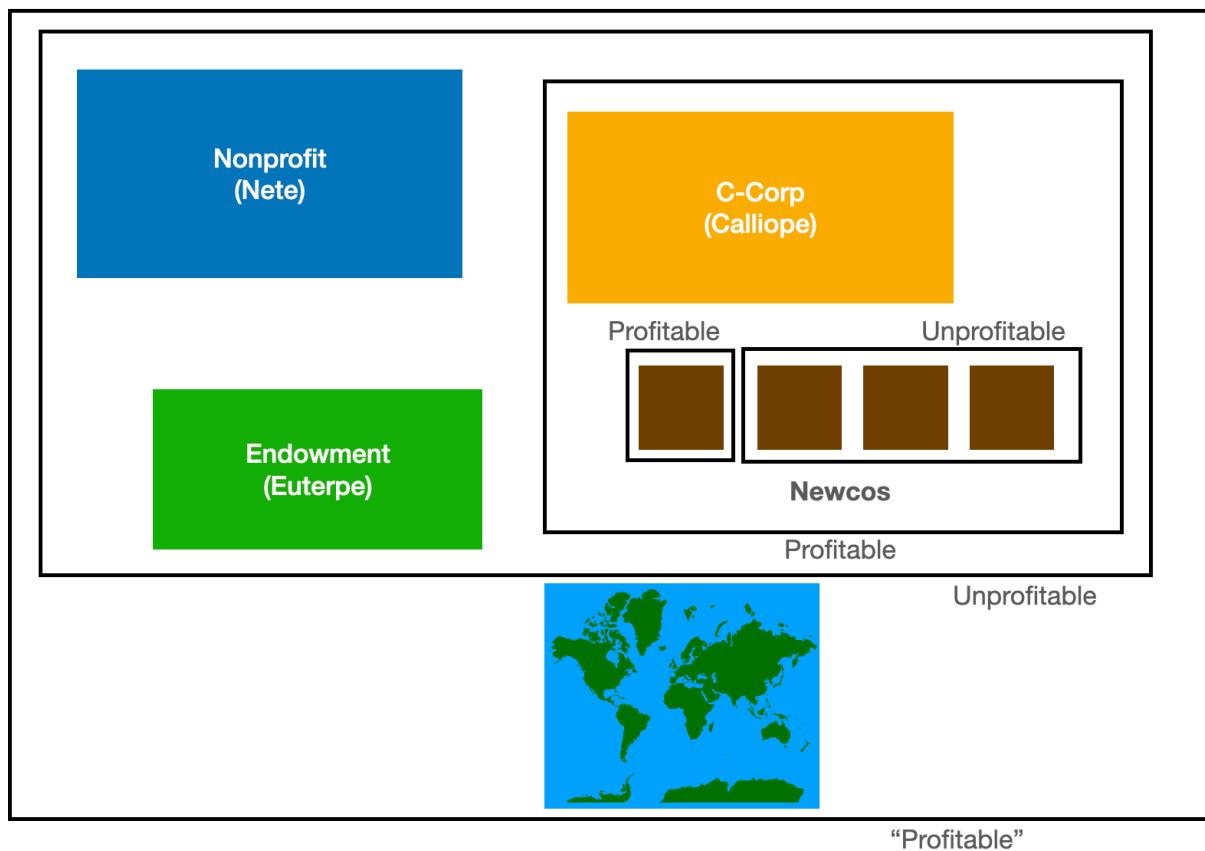
The points system would also enable people to participate in multiple ways instead of different types of shares shunting people into different buckets according to their roles; there would be no accounting hassle for an employee who also wanted to invest. There's also a psychological component in the difference between adding a grain to a growing pile of sand and getting a slice of a fixed pie. I realize this all comes with a whiff of utopian faeries and rainbows — regulatory constraints and investor pressure will undoubtedly make this full system infeasible. It's important to lay it out however, so that we can get as close as possible without going to jail.

The trickiest pieces of the puzzle are the questions of who does the work, who owns it, and who employs people. It would be awkward if Nete had a bunch of employees who just happened to get compensated with a stake in Calliope. A tentatively feasible way to separate Nete and Calliope is as follows: all of the PMs are Calliope's employees. They are then contracted to Nete, who owns any resulting work. If the work gets to the point where a company makes sense, Nete then can license that work at fair market value to Calliope who would then spin out companies. At least in theory, this arrangement allows PMs to have an equity stake while at the same time making sure that Nete, not Calliope (and therefore profit) are driving research agendas.

Calliope's existence won't free Nete from depending on donations to fund

²⁵⁴ As with any metric, it will be imperfect. It's not actually possible to convert between four years of service and \$200k. Both are essential and uninterchangable.

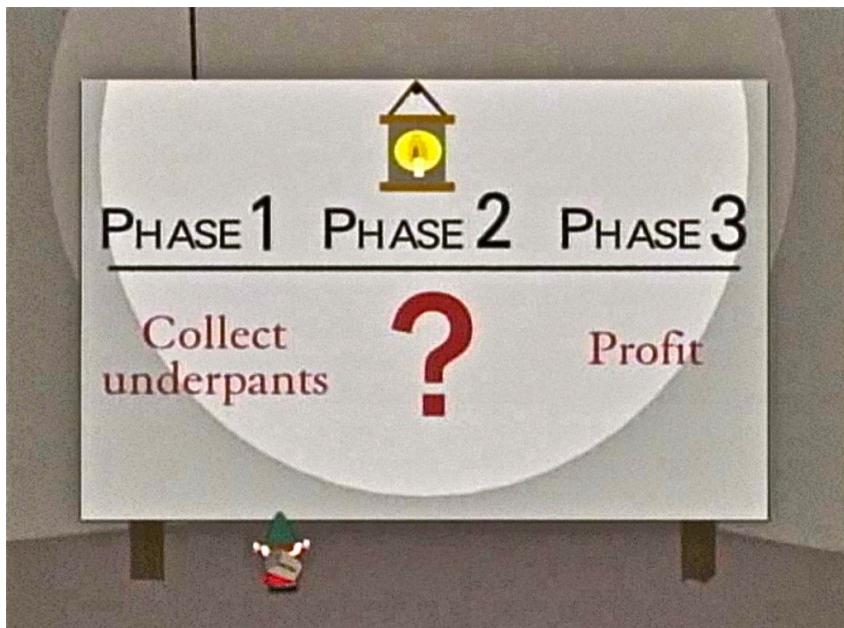
operations for many years. However, Calliope *can* possibly free Nete from depending on donations in the long run by filling Euterpe's coffers. This scheme relies on two big assumptions — that Calliope will eventually become very valuable and/or that Nete will be able to get enough donations to cover operations that, through some combination of those donations and money from Calliope, there will be enough to slowly build up Euterpe's principal to the point where its dividends can fund Nete indefinitely.



PARPA will Go Through Three Phases of Organizational Growth

Organizations tend to resemble punctuated equilibri²⁵⁵a — instead of always changing at a continuous rate, they tend to go through periods of relative stability punctuated by ‘phase shifts’ when they hit a critical level of resources. In other words, organizations evolve more like Pokémon than (growing) people.

Most organizations go through different forms but rarely make them explicit. Instead, the situation resembles the South Park Underpants Gnome’s business plan: there’s an initial form (collect underpants) and a final form (profit) with a big question mark in between them. It’s hard to evaluate how good such a plan is from the outside but perhaps more importantly, it’s hard to use a gnomish plan internally — when should you push to the next phase? When should you exert discipline and remain in the current phase because you haven’t met key conditions for a phase shift? Making the different forms explicit can get everybody on the same page. As with so many other things, making forms legible also makes them vulnerable — both to criticism and to accusations of hypocrisy if they end up not playing out as predicted (which is almost inevitable). In my calculus the clarity and intellectual honesty is worth the risk.



²⁵⁵ Evolutionary systems (may) go through long periods of slow change punctuated by periods of rapid change. See Punctuated equilibria: an alternative to phyletic gradualism

Through this lens, we can think about PARPA as taking on different ‘forms’ over time. Each form will require different resources, focus on a different set of activities, and generate different outputs. These forms roughly correspond to each of the major hypotheses on PARPA’s path through the idea maze.



PARPA could use tranches to walk the line between long-term funding and step-by-step derisking

One pragmatic reason to explicitly divide PARPA’s development into discrete forms is to enable tranches for both donors and investors. As I’ve noted several times, there are many tensions between starting small and the other pieces of the DARPA Model. Tranched donations and funding could be one way to relax this tension.²⁵⁶

Milestone-based funding is common in the pharmaceutical world, where developing a drug takes many years and potentially billions of dollars, yet

²⁵⁶ For some more complex and well-thought-out tranched funding schemes, see Funding Long Shots.

follows a fairly predictable trajectory through the different stages of FDA approval. The predictability and structure do not reduce the risk associated with the venture, but they do reduce the uncertainty about possible outcomes.

PARPA could conceivably raise money in a similar way — soliciting small donations that are automatically followed up by larger ones once specific milestones are hit. This approach would allow the organization to act as though it had secured long-term funding without being under the pressure to deploy a large amount of money. Milestone-based funding would also allow donors and funders to hold onto their money until the milestones were hit and feel like the organization was derisked before deploying larger amounts of money.

One danger is that a donor would choose to renege on their commitment once a milestone is hit. This unfortunately happens with capital calls in funds but hopefully the chances of it happening in this case could be small both because hitting milestones will be less unexpected than a capital call (which could come at any time) and clear milestones could enable a firmer contract.

Creating and agreeing on milestones will be a challenge. Milestone-based tranches requires clear milestones. One of the reason it works so well in medical technology and drugs is that passing FDA trials are unambiguous milestones that everybody can agree on. PARPA won't have the luxury(?) of a staged government gatekeeper as the primary barrier to success so we'll need to select milestones carefully and do the work to get donors to buy into them. My intention is that laying out the forms below is a first step in that process!

Form 1: Program Design

- People: 1-5
- Budget: 0-\$500K/year
- Outputs: Program designs



In this form, PARPA will be just one or two full-time individuals designing possible programs. Remember, it is important for big things to start small! These PMs will be focused on an iterated process of ‘hub-and-spoke’ style work where an PMs reads many papers and talks to many people to form their own hypotheses and couples that with many-to-many workshops where several experts come together to hash out possibilities.

In this form, the organization’s primary output will be artifacts — let’s call them roadmaps. These living documents will work backwards from a clear goal to the precise work that would maximize your chance of getting to that goal; call out who might be best suited to do that work (there are many pieces of equipment or tacit knowledge that only exist in one or two places in the world!); and explain how to think about the work in terms of timelines, branch points, and milestones. The roadmaps will inevitably need a few iterations of real experiments to ‘derisk’ key assumptions. Keep in mind that a roadmap doesn’t mean that a program is ‘shovel-ready.’ There will be a lot of work to get all the pieces in place eg. figuring out who will do the work, signing contracts, sorting out equipment, etc.

Ideally, other organizations would act on these roadmaps but I wouldn't bet on it.

Ideally we will evolve out of this form as quickly as possible once we have a good handle on the hypothesis that it's possible to design programs in a systematic (and therefore repeatable) way.

There is no critical mass to get to this stage but the critical mass to get out of it is to have designed around five programs that are worth moving forward with.²⁵⁷ We need to generate a sufficient number of roadmaps to both show both that the planning process is repeatable and that there are enough areas in PARPA's early sweet spot to make this whole endeavor worthwhile.

The budget for this stage is just enough to cover individual salaries and the costs of running workshops — renting out a cabin in the woods and (hopefully) food and transportation for people, so a few thousand \$\$. The tricky bit here is that our ability to dig into problems and possibilities deeply enough to do good program design may be coupled to our ability to execute on those programs. In my experience²⁵⁸ people are much more willing to talk shop about precise potentials when they can see the conversation leading to an actual program.

If other organizations do want to act on the roadmaps, there is perhaps a form 1.5 in which the boundaries between PARPA and the executing organizations becomes fuzzier. We could bring representatives from those organizations into the process to increase the chance that the output becomes reality. Similarly PARPA people could consult with organizations executing on roadmaps to keep the roadmaps alive and make sure nothing is lost in translation.

²⁵⁷ While a DARPA-riff will need at least seven programs to demonstrate viability, we could design the remaining programs once we've moved to Form 2.

²⁵⁸ I've received drastically different response rates sending emails with the slightly different wording "potential privately funded research program ..." and just "potential research program ..."

Form 2: Architecting Institute

- People: 5-20
- Budget: \$10M/year
- Outputs: Self-sufficient but temporary research organizations



In this form PARPA will comprise several program managers working on program design but expanding beyond Form 1's process by running seedling experiments and getting programs off the ground. The organization's primary output will be (semi) independent organizations along the lines of Focused Research Organizations²⁵⁹ whose mission is to execute on a program.

Once PM has driven a program to a point where they have good answers to all

²⁵⁹ See Focused Research Organizations to Accelerate Science, Technology, and Medicine, Again.

the questions in the Heilmeier Catechism²⁶⁰ or an equivalent ‘gate’²⁶¹ they would need to get all the pieces in place for the full program organization — funding, incorporation, research contracts, and hiring — and either move to the program organization full-time or transition leadership to someone else.

The funding for these program-executing organizations could come from a myriad of sources — the government, philanthropists, or investors depending on the nature of work to be done. Frankly, this transition from a loose set of connected projects to its own organization seems incredibly tricky to do well. There are many disconnected pieces that need to come together. Compromises to get funding or convince people to join could end up derailing the program. PARPA should be able to help the creation of these temporary organizations by maintaining connections to funding sources, labs, and contract research organizations, providing initial space, and generally honing process knowledge. It might be possible to get around these problems by building out the idea of Exploratory Program Organizations that enable a program to undergo a more continuous transition to an independent entity.

There is a minimum number of programs that PARPA needs to be working on in parallel — probably in the five to seven range. If the programs are actually high risk, any single one is more likely to fail than not, so if there were only a few programs, the chances that all of them would go nowhere and take down the whole organization with them would be quite high. If each program has a 5-10% chance of success (as is the case in DARPA²⁶²) then you need to run seven programs to get a 50% chance of seeing one of them succeed.

A critical mass of programs needs a critical mass of program managers to run them. In addition to the bare minimum number of people needed to run the programs, there’s some number above which the group starts passively generating more ideas than all the individuals on their own. This number is

²⁶⁰ See <https://www.darpa.mil/work-with-us/heilmeier-catechism>

²⁶¹ Speculatively, it might be possible to improve on the Heilmeier Catechism with something more quantized - similar to Donald Braben’s BRAVERI scale.

²⁶² See DARPA - Enabling Technical Innovation

probably around five to seven: a dinner party's worth.²⁶³

This form is where the director emerges as a separate role to focus on the organization itself. The director wasn't even worth mentioning in Form one because in the beginning the director will be basically a PM with some extra administrative work.

Estimating the minimum level for the budget in Form 2 is tricky. Too low and you're unable to do the work to show that the whole organization has potential. DARPA seedling programs have total budgets between \$1-5M. If we assume that you want to complete the program in 6 to 18 months, that's a rate of \$0.7 to \$10M per program per year. If you need a critical mass of five programs at any one time, that's \$3.5 to \$50M per year in research costs. Obviously this is a huge range drawn from multiplying the extreme ends of the time/budget spectrum. A more reasonable estimate would probably be \$10M/year. This number feels like it's skirting the edge of setting the organization up to fail because of the inevitable pressure to deliver results ASAP. So in total, the critical budget to consider moving to this stage is around ~\$10M/year but hopefully higher.

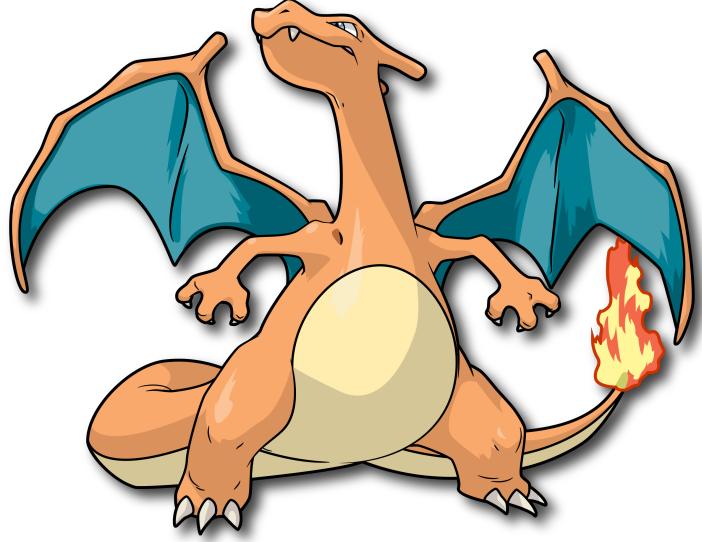
While the \$10M/year number is tiny compared to the amount that the US government spends on research, it is massive compared to most relatively new organizations. Donors might be understandably hesitant to commit to that amount for the several years you would need to see real outputs. Ideally verdicts would be withheld until after the created semi-independent programs had a chance to show results. This is a place where tranches could be clutch.

Form 3: An Institute for Creative Technology

- People: 20 - 100

²⁶³ You also need some amount of slack for these benefits to kick in. If everybody is spending 110% of their time focused on only their own programs any non-program-related interaction is mentally expensive.

- Budget: \$60M-\$3B/year
- Outputs: Open research, companies



This phase is the fully-evolved version of PARPA that can hopefully exist productively for a very long time. In this phase, PARPA will not only design and derisk programs, but will run them as well.

Running programs without creating and funding an entire new organization has several advantages. Internal programs enable higher-risk programs because you can shift money away from a failed program to a healthy one. Internal programs smooth the transition from “hey we should do this!” to actually doing it — you don’t have to go and raise a bunch of money, transfer knowledge and spin up a whole organization. Instead, all of that can happen organically over the course of the program.

In this phase the organization’s outputs are whatever the programs turn into — whether it’s startup companies, research nonprofits, or knowledge and IP that we work to diffuse into existing organizations. PARPA will still need to do a lot of work to make sure that the programs transition to whatever form will maximize their impact. Everything else about the organization will be a fairly natural extension from Form 2.

Like Form 2, Form 3 does need a critical mass of simultaneous programs because

of the program's high chances of failure. This number is probably around seven, based on the back-of-the-envelop exercise that a 10% success rate requires seven programs in order to give the organization a 50% chance of a single successful program. Assuming each of seven programs costs ~\$8M/year and you're continuing to do seedling programs as well (per Form 2) you wind up with a minimum budget of ~\$60m/year. Interestingly, this number roughly corresponds to the original ARPA IPTO budget that midwifed the personal computer (\$2020 47M).

Eventually, we aim to get to the scale of DARPA: ~100 programs with program managers deploying a budget of ~\$3B. Of course, that steady state can and should be approached incrementally after many years of operation. I realize this target is incredibly ambitious and there is a large chance we will never get there!

Potential Programs

I struggle with how far down the ladder of abstraction to climb when talking about specific programs. On the one hand, I have built a few strong hypotheses and having examples is always helpful for what is otherwise a very abstract proposal. On the other hand, readers (not *you* of course, but other people) tend to over-focus on those examples to the point where questions and disagreements over the examples dominate everything else. ("Forget all the model, forget the plans, why are you on *that*?") Which programs we'll work on is both one of the most important and least fixed piece of the plan. But I've already asserted that precise visions are more likely to happen so I will hold myself to my word. Keep in mind that these are hunches that need a lot of work to even verify that they're worth creating a program around. As it says on open house furniture: for display only (but perhaps available later). In this section, I'll present both a list of hypotheses that may be worth designing programs around and two more detailed descriptions of programs I suspect are particularly promising.

Loose Program Hypotheses

A caveat: you are probably going to think that most of these examples are objectionable for one reason or another. Some will sound like fantastic ramblings, others will sound too eminently practicable to need new institutional support.

- Reconfigurable black-box factories are a systems and robotics problem that could drastically change what gets made where.
- 3D printing could potentially enable soft robotics to become a useful thing.
- Non-human-centric biology could unlock new paradigms of physical technology. For example, you could make high strength to weight linear actuators that effectively run on garbage if you could create an artificial muscle coupled to an artificial metabolism.
- Multi-purpose farming robotics could change the way we grow food, especially indoors.
- New simulation paradigms — especially hierarchical systems of simulations — could enable everything from new materials to generative design to new generations of robotics.
- ‘General purpose’ telerobotics is a systems engineer problem that could both address global problems and enable people to do things they never could before.
- “Positional chemistry” driven by some combination of protein 3D printers, nanomachines embedded in metal-organic frameworks, and other tools could begin to deliver on the vision of atomically precise manufacturing.

General Purpose Telerobotics

Possibilities

One of the most iconic scenes in *The Matrix* occurs soon after Neo has escaped The Matrix for the first time and is learning that you can download in-Matrix skills directly into his brain. He plugs a cable into the back of his neck, thrashes around, opens his eyes and announces “I know kung fu.” It has become shorthand for the idea of ‘downloading’ an ability. While telerobotic technology cannot give people superpowers directly, it can create many of the outcomes. There is something intangibly beautiful about giving people more agency. In the same way that computers can be a ‘bicycle for the mind’ enhancing our creativity and thinking abilities, robots could be a ‘computer for the hands.’

The first and most obvious power that telerobotics can give people is the ability to teleport. There are obvious benefits (less commuting for physical jobs, an expert chef being able to cook for people around the world, etc.) but one of the less obvious ones is the ability to fully utilize relatively rare pieces of equipment. At 2am, people on the other side of the world can continue research work.

Teleoperated robots could be the size of buildings. In addition to fulfilling everyone’s fantasies of commanding giant mechs, massive remotely-operated robots could do the work of large crews on construction sites or during emergencies. Both are high-uncertainty situations where we won’t be trusting fully-automated robots anytime soon. On the flip side, teleoperated robots could be small enough to go inside a human body, crawl inside walls, and generally operate in places a person could never access non-destructively. Robots the size of pills could make *Fantastic Journey*a (less adventurous) reality. To some extent this already exists in the tools for laparoscopic surgery and the DaVinci robot. However, we should see these as crude hints of what is possible. Robots can have a plethora of form factors. Researchers have worked on everything from snake robots to gecko robots to octopus robots. Telerobotic technology can enable people utilize the advantages of these form factors like any shape-shifting superhero.

Even at human scale, robots can have physical abilities far beyond humans.

Electricity-powered robots don't need air. Robotic hands can hold things that are much hotter or colder than a human. Powerful motors can enable a robot (and by extension its operator) to have superhuman strength, while precise stepper motors can enable people to act with steady precision that few human hands can. Regardless of their main powers, superheroes are generally much more damage resistant than a normal person. A remotely operated robot enables squishy humans to approximate this damage resistance. There's also the fact that it's much more ethical to put a robot in danger than a person. For now, robots are so expensive and telerobotic technology is clunky enough that it's worthwhile to pay the liability insurance and send people into danger. Dedicated work on telerobotic technologies can change this 'fact.'

Challenges

If telerobotics is so great and not just a pipe dream, why are real telerobotics platforms basically just iPads on Segways? A big part of the answer is that there are many different pieces that need work in the context of a system. A general purpose telerobot needs to enable a person to smoothly interact with an environment.

Without laying down telerobotics-dedicated cable, telerobots will always need to operate over the internet. That means dropped packets and variable lag. While these are slightly annoying during a video call, they can be disastrous for a delicate task. Naïvely implemented, haptic feedback over a laggy connection can create a self-reinforcing loop that could (without force limiters) rip off an arm. There are many potential ways to get around lag: onboard autonomy that can translate 'rotate this', 'pick that up' into lower-level path planning; generating a model of the robot's environment that a controller interacts with and then the system translates into robot movement on the other end; or the less technical approach of cutting a deal with a telecom company. The trick is that how well each of these approaches 'works' depends on other pieces of the system.

Human touch is pretty extraordinary — pick up a pen if one's handy (any

small object will do), read the next sentence, and then close your eyes. Once your eyes are closed, twiddle the pen in a few loops, bring it back to its original position, and then open your eyes again. The extraordinary thing is how *easy* that probably was. Most modern robotics and telerobotics operates primarily through visual sensors — either passive, like a CCD or active like LIDAR. Compared to gigapixel visual sensors, modern tactile sensors are primitive. Human skin can have neurite densities in the thousands per square mm²⁶⁴ while the cutting edge BioTac tactile sensor measures forces via the differences between 19 electrodes.²⁶⁵ The situation isn't much better interfacing with the human hand on the other end. One upshot might be that the binding constraint on telerobotics is haptics technology. Perhaps! But at a system level there may be ways to achieve better performance even with low-quality tactile sensors: multiple cameras, simulated tactile responses, [Project Soli](#)-style radar-on-a-chip, or autonomy in the loop.

Human hands are also versatile — those same hands that can thread a needle or hold a baby can also lift hundreds of pounds or smash a brick. Robotic actuators (the fancy word for ‘things that do things’) are nowhere near that good — either they’re too weak to do any damage or liable to destroy things without very careful control. There are so many potential ways to improve this situation: from haptic feedback loops, to mechanically compliant arms, to soft grippers. Artificial muscles, pneumatic systems, and pulley-driven arms/hands are all potential ways to approach the versatility of human limbs and hands. The trick is (and I will sound like a broken record here) that the ‘best’ solution only exists in the context of a system — if you have lower lag, you need to worry less about arm compliance, if you have a four-armed robot you could switch between strong arms and delicate ones, etc.

So far we’ve only talked about the technical challenges. There are many

²⁶⁴ See The density of remaining nerve endings in human skin with and without postherpetic neuralgia after shingles

²⁶⁵ See The BioTac - Multimodal Tactile Sensor

people-based constraints on telerobotic development. Telerobotics has become seen as the red-headed step child next to work on full autonomy. It's seen as nothing but a stepping stone to a near-future inevitability. Hopefully previous sections at least nudged your belief that such assertions are like saying that work on productivity software is useless because it will all be automated anyway. The work that *is* happening is rarely done in a real context-of-use outside of a few exceptions; researchers usually demonstrate different subsystems on benchmark tasks, publish a paper, and call it a day. These benchmarks are barely even comparable because every telerobotics lab cobbles together their own unique setup from different commercially available components.

Actions

What would a PARPA program to go after general purpose telerobotics look like?

One concrete possibility is to work with academic and commercial groups to create a standardized telerobotic research platform. Both the [PR2](#) and [Baxter](#) robots simultaneously spurred a lot of robotics work by providing an extensible research platform and were commercial failures. PARPA, in its role of taking on work that both academia and startups won't touch, could coordinate work to build an extensible telerobot and the work to figure out how to make it a maximally useful general tool.

Another possibility might look like picking one extremely hard but obviously important context-of-use and then coordinating work to build a prototype system within that context. Hopefully, a challenging context-of-use with many contradictory requirements would prevent the system from overspecializing. The goal would be twofold: first we would try to get the system to a point where commercial entities see enough potential to move the ball forward. Second, you could imagine reaching a place where improvement is a matter of optimizing individual components, rather than the system as a whole. Elder care is the most salient context-of-use; it's important and involves everything from big strenuous

(but delicate) tasks like physically helping someone move to small precise things like preparing food.

Positional Chemistry

Possibilities

I'll stand on the shoulders of some giants:

Up to now, we have been content to dig in the ground to find minerals. We heat them and we do things on a large scale with them, and we hope to get a pure substance with just so much impurity, and so on. But we must always accept some atomic arrangement that nature gives us. We haven't got anything, say, with a "checkerboard" arrangement, with the impurity atoms exactly arranged 1,000 angstroms apart, or in some other particular pattern.

What could we do with layered structures with just the right layers? What would the properties of materials be if we could really arrange the atoms the way we want them? They would be very interesting to investigate theoretically. I can't see exactly what would happen, but I can hardly doubt that when we have some *control* of the arrangement of things on a small scale we will get an enormously greater range of possible properties that substances can have, and of different things that we can do.

Consider, for example, a piece of material in which we make little coils and condensers (or their solid state analogs) 1,000 or 10,000 angstroms in a circuit, one right next to the other, over a large area, with little antennas sticking out at the other end – a whole series of circuits. Is it possible, for example, to emit light from a whole set of antennas, like we emit radio waves from an organized set of antennas to beam the radio programs to Europe? The same thing would be to *beam* the light out in a definite direction with very high intensity. (Perhaps such a beam is not very useful technically or economically.)

...

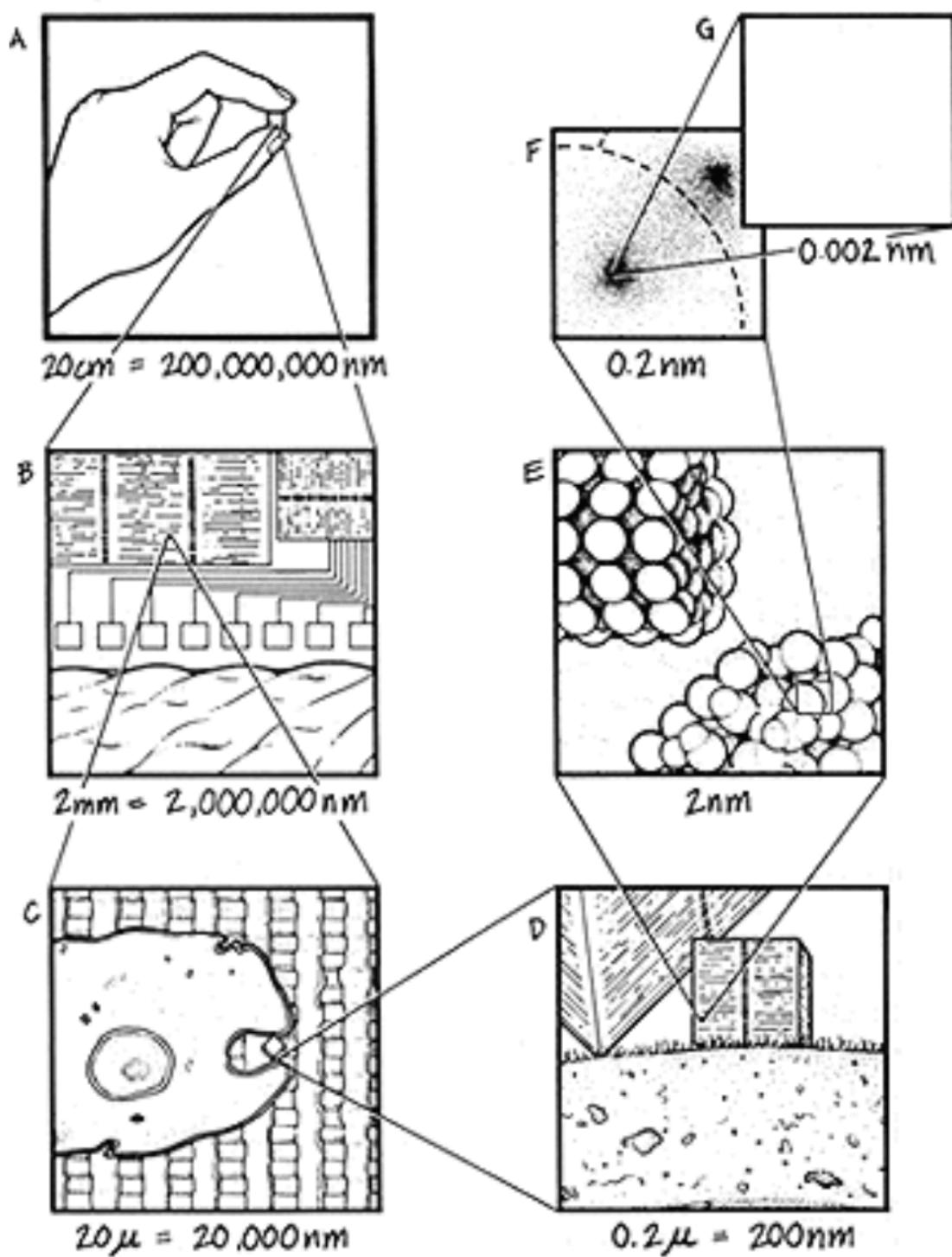
But it is interesting that it would be, in principle, possible (I think) for a physicist to synthesize any chemical substance that the chemist writes down. Give the orders and the physicist synthesizes it. How? Put the atoms down where the chemist says, and so you make the substance. The problems of chemistry and biology can be greatly helped if our ability to see what we are doing, and to do things on an atomic level, is ultimately

developed – a development which I think cannot be avoided.

— Richard Feynman [There's Plenty of Room at the Bottom](#)

A short summary of what molecular nanotechnology will mean is *thorough and inexpensive control of the structure of matter*. Pollution, physical disease, and material poverty all stem from poor control of the structure of matter. Strip mines, clear-cutting, refineries, paper mills, and oil wells are some of the crude, twentieth-century technologies that will be replaced. Dental drills and toxic chemotherapies are others.

— K. Eric Drexler [Unbounding the Future](#)



In short, successful positional chemistry would allow us to turn the arrangement of atoms from an indirect, stochastic process to a direct, designed process. Biology already does this — you could think of the relationship between positional chemistry and biological processes like the action of ribosomes or

kinesin as similar to the relationship between aircraft and birds. Unfortunately we're still lashing wings to our arms and jumping off churches.

Challenges

You would need an entire book to do justice to the challenges facing anyone who wants to make the Feynman/Drexler vision a reality. However, it boils down to the fact that atoms aren't just little balls you can stick to each other. The atomic-scale world effectively has a different set of rules and demands almost entirely new engineering paradigms.

The world on the scale of atoms works very differently than our macro-world. (As an intuition pump, a carbon atom is ~0.3nm across, a water molecule is ~0.27nm and DNA is ~2nm wide.) Gravity doesn't matter, everything is floppy and constantly jiggling, and quantum effects are a going concern. Some people argue²⁶⁶ that these attributes make it impossible to do anything at this scale besides 'stochastic building' — effectively putting a bunch of ingredients together, shaking, and counting on their properties to direct the process and output. While I'm optimistic that the properties of the nanoscale world don't make positional chemistry impossible, it *will* require drastically new ways of doing engineering.

I'll use the example of a DNA '3D printer' to illustrate some more general challenges in positional chemistry.²⁶⁷ Imagine a DNA-based protein 3D printer / pick and place machine — a precisely positioned write head that can pick up and precisely place prefabricated proteins on a work surface. At some point any positional chemistry system need to interface between stochastic processes and more deterministic ones at the same scale. The printer might need to change 'tools' (antibody-like proteins?) by flowing a high concentration solution of the new tools over the device. These tools will be constrained (for a long time) by the

²⁶⁶ See the Drexler-Smalley debate for strong arguments against positional chemistry.

²⁶⁷ Note that I'm *not* saying this is necessarily the right thing to build, but it does seem promising.

necessity to be fabricate them through traditional methods eg. putting stuff in a tube and shaking. The same goes for the protein ‘lego bricks’ that the printer is arranging. The bricks will need to fold the good old fashioned way, so they’ll be limited by the shapes and capabilities of amino acids (which luckily are provably capable of creating everything from bone to wood to shell). The upshot is that this will in no way be a ‘universal matter printer.’

These constraints will cascade up to what you can build and the steps that you need to take to build it in ways that traditional engineering has no good way of wrapping its collective mind around. To make things even harder, the traditional engineering tools of models and simulations basically need to be reinvented for this new world. While the topics themselves still matter just as much (or more!) you would need to throw out your college textbooks on thermodynamics, statics, and dynamics. Great progress has been made on simulating processes on the nanoscale, but there is a lot of work to be done both on making the simulations more versatile and on using them as an engineering tool.

And how do you know that your simulations were accurate? Observability is a challenge both for model development and for positional chemistry in general. Unlike a macro-scale system, you usually can’t just look at a nanoscale system and see whether it looks like you expect it to based on the model. While scanning electron microscopes, atomic force microscopes, X-ray crystallography, and other tools make the nanoscale world not entirely inaccessible, they’re severely limited in the conditions they can observe and the data they produce often requires interpretation. In many situations, you’re limited to observing secondary effects to tell you whether things are working the way you expect. These indirect observations make it harder to do failure analysis and increase the length of feedback loops which is yet another challenge.

You also need to tell the printer what to do! Unlike a macro scale printer, you can’t just hook up some wires to a stepper motor or turn a crank on a mill.

Instead the current toolkit is limited to some combination of light, chemical gradients, and ‘preprogrammed’ cascading reactions.

A meta-challenge is that there isn’t a clear ‘best system’ for achieving positional chemistry. I’ve been using a DNA-based 3D printer as an example but maybe the printer should be built out of proteins, DNA-cast metal, or implosion fabricated materials. Or the 3D-printer approach might be the incorrect first step. An entirely different, but potentially exciting approach is to embed molecular motors and switches in a metal-organic framework scaffolding with the whole system acting as a tiny ‘assembly line’ with each step adding or changing the ultimate product. Each approach has its own specialists and advocates. Maybe the most effective approach involves some of all of them. It’s a complex systems-engineering question that doesn’t yet allow you to abstract things into black boxes in the way that systems engineers like to operate.

Arguably, the biggest meta-challenge is the human one. Everything that I’ve mentioned so far is primarily in the domain of people who are part of academic chemistry or biology departments. Novelty and papers are strong incentive for them. The answer to the question “this work is excellent — how do you think it could be integrated into a bigger system?” is often “No clue. Not my department.” As we’ve discussed, novelty and papers are not bad but they can be at odds with engineering design. In addition to being fragmented across different academic disciplines, this entire area is heavily stained with what I would call ‘hype-fallout.’ During the 90’s people became very excited about the possibilities of ‘nanotechnology.’ As a result, the government started pouring money into the area in part by pulling funding from other areas. This funding shift pushed anybody who was doing anything remotely related to processes that occurred on the scale of ~1nm to relabel their work as nanotechnology, diluting the term, and blunting progress towards ‘atomically precise manufacturing’ — another radioactive term. Combine discipline dilution with overpromises and a healthy dose of crankery and the modern hesitance to take the positional chemistry as an engineering problem seriously is entirely reasonable.

Actions

In the near term, a PARPA program could set out to answer the question “is it possible to build a system that does positional chemistry?” Or perhaps more aggressively “is it possible to build a system that does reconfigurable positional chemistry?” The initial steps would be to fund a number of parallel efforts with the explicit intention of figuring out how to integrate them should they prove promising. Despite the grim picture the challenges paint there are some reasons to be hopeful. In the past decade there’s been an explosion of tools and techniques to design and fabricate roughly-deterministic structures out of biological materials — especially DNA — and do useful things with them. There are many new ways to simulate and synthesize useful proteins. This work might provide a jumping off point for positional chemistry.

A non-exclusive list of parallel efforts might look like:

- A project to develop precise actuators for a DNA-based printer. Perhaps an expansion on William Shih’s work.²⁶⁸
- A project to develop protein building blocks (probably based on coiled-coil proteins²⁶⁹) that could be used in a DNA-based printer.
- A project to simulate the protein building blocks using machine learning — perhaps a system that uses something like [Alphafold](#) in a reinforcement loop.
- A project to simulate what could be built *with* the protein building blocks.
- A project to develop different “tooling” for the 3D printer — perhaps structures derived from antibodies that have a similar ability to bind to specific sites on the protein building blocks.
- A project to develop artificial amino acids that could be part of the building blocks and bring non-organic molecules into the system’s output.

²⁶⁸ See DNA Strand Displacement-Driven DNA Origami Tools and Materials.

²⁶⁹ See Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments.

- A project to poke at how Metal-Organic Frameworks²⁷⁰ might play with the other projects.
- A project to look at how DNA/protein metal ‘casting’²⁷¹ might play with the other projects.

Reasons why an PARPA would succeed or fail

A theory which is not reputable by any conceivable event is non-scientific.

— Karl Popper, Science as Falsification

The ideas I’ve laid out are not scientific, but it’s still important to ask “How will we know that there is merit to these ideas? What evidence would suggest that they’re wrong?” It’s tricky because any outcome will doubtless have many causes. Complicating things further is the fact that it’s not always clear whether an outcome counts as a success or a failure — does dying in the process of killing the minotaur count as success or failure? Grappling with these questions explicitly before the fact is unusual but can hopefully be a bit of a golden thread — giving us a sense of when we’re moving in the right direction or straying into a trap. I hope that this exercise also enables PARPA to be a useful case study, regardless of its outcome.

Organizational failure scenarios are absurdly contingent, especially when traversing less-trodden paths in obscure mazes. Sometimes it’s not a monster that gets the hero, but a random falling stalactite. The potentially large effects of small unanticipated factors is a big reason for this entire piece and this section in particular; it’s important to record hypotheses and how they evolve over time before narrative-building inevitably kicks in around either success or failure. Obviously there are many unknown unknowns, but if I’ve done a reasonable job,

²⁷⁰ See Ultrafast rotation in an amphidynamic crystalline metal organic framework.

²⁷¹ See Casting inorganic structures with DNA molds.

this section covers most possible scenarios.

Hints of Success

PARPA's long-run success looks like being a contingent cause of humanity becoming more awesome. Admittedly, it's hard to get more nebulous, long term, and debatable than that. Instead of a well-defined success condition, it's probably more useful to paint several concrete and shorter-term possibilities that would suggest that we're moving in the direction of success.

Uncomfortably, any one of these success indicators in isolation probably does not point towards broader success. However, it is important to list them in an attempt to minimize either positive or negative hindsight bias. Looking at organizations in hindsight is generally just narrative-building full of both too much "well this was actually a success because..." and too much "that didn't really do anything..." So while any isolated item on this list doesn't count as success, together they're directionally suggestive: the more boxes the organization can check off, the more it will look like a success.

- *Shift ideas from outlandish to mundane.* Technological paradigm shifts don't come out of the blue. They are usually preceded by shifting Overton windows around what is impossible, what is feasible, and what is expected. These opinion shifts aren't necessarily causal so shifting opinion should not be a goal in and of itself but could be a useful leading indicator of doing good work.²⁷²

- *Produce useful ideas or artifacts that other organizations pick up and run with.*

Creating convincing programs, prototypes, or system-seeds that other organizations pick up and run with is absolutely a leading success indicator. Andy Matuschak put it well: the goal is "developing ideas far enough that they become "obvious," the banal fodder for half a dozen companies in a future YC

²⁷² This distinction between leading indicator and goal in of itself is of course the root of Goodhart's law. We can hopefully sidestep the trap of leading indicators becoming goals by having too many indicators to pursue all of them without just pursuing the thing they indicate!

batch.”²⁷³ There are traps to be aware of here as well — in the same way that an acquired company’s product is often quietly shut down, it’s a real possibility that other organizations can claim to be pushing ideas or technology forward while quietly smothering it or completely mangling it.

- *Create a large impact in a small niche.* On the scale of a few years, it’s impossible to create unambiguously impactful innovations. Non-digital dissemination is always slower than you expect. However, it may be feasible to create big outcomes among a small group of people — perhaps a research domain or a specific set of practitioners.

The last two success indicators are perhaps not leading indicators but ‘alternative win conditions.’

- *Serve as a jumping-off point for eventually extremely impactful people.* History is full of largely-forgotten ‘scenius’²⁷⁴ that quietly lie in the past of an uncanny number of impactful individuals. Creating even a short-lived scenius, like the Rad Lab or Willow Garage would be a worthy result.
- *In death, serve as an inspiration for similar models.* A single organization can’t fill all the gaps in the innovation ecosystem so generating lessons and impetus for other organizations is an actual win (as opposed to a bitter consolation prize). Of course, it’s hard to be inspirational when you haven’t produced anything of note. Shooting for inspiration without more tangible outcomes is the path of people who have ‘visionary’ on their business cards. So while it’s possible for the organization to die prematurely, it still needs other successes.

Longevity and notoriety are not success indicators

The ambiguity between success and failure is compounded by organizational longevity. Any innovation organization that survives long enough will usually end up with one or two wins to its name. I wouldn’t call the NSF and NIH shining examples of success but they have supported many paradigm shifts and

²⁷³ See Reflections on 2020 as an independent researcher

²⁷⁴ See Scenius, or Communal Genius

Nobel prizes. I would argue that they achieved these wins through their sheer size and longevity. Perhaps there's a useful efficiency metric like "Nobel prizes per \$" but I have no clue what it is. The paired realities that solutions R&D doesn't lend itself to Nobel Prizes and the number of intermediate steps and different actors to get to a clear win makes an efficiency metric hard.

There is something to be said for simply surviving long enough to find the right project or hop on the right S-curve. Innovation organizations can take a long time to hit their stride. Bell Labs didn't start producing the work we laud them for today until more than a decade after it was founded. If for some reason it had closed its doors in 1935 (10 years after it was started) it would be barely a blip on the historical radar. Similarly, ARPA's first programs were at-the-time critical but now-forgettable, pursuing questions like "can we stop a missile strike by exploding a nuclear device in the upper atmosphere?"

The ambiguity around intermediate organizations' impact and the legitimate value of longevity make it easy to tell yourself and the world that you're building up to something great while in fact you will just limp along forever. There are many organizations we can both think of that have done this.

For PARPA, raw survival is not success. One could imagine us surviving for a more than a decade, putting out white-papers perhaps backed up by little proof-of-concept experiments without any significant impact. Even if PARPA becomes well-known, notoriety doesn't count as success unless it stems from enabling outputs that wouldn't have happened otherwise or serving as a direct model for other people to do that work. At the same time, even if produces impressive outputs, it's a failure if PARPA winds up effectively indistinguishable from a consulting firm, a government contractor, a startups studio, or other possibilities. I realize this is a high bar.

What would failure Look Like and Why Would it Happen?

The line between success and failure is more nebulous than we'd like to admit. There's an obvious difference between abject failure and clear success but a rich spectrum lies between them. This ambiguity is especially prominent for

organizations like PARPA that play intermediate roles in the relay race of turning ideas into impactful things in the world. We stand on the shoulders of too many giants to give them all credit.

Precise visions of failure are important because post-hoc success can be woven into a failure narrative and failure can be woven into a success narrative. Some organizations fail to achieve their explicit goals but should go in the success column. Taylor-Era PARC²⁷⁵ and Willow Garage were not particularly long lived (12 and 7 years respectively) nor did they live up to the expectations of their creators. However, they ultimately spawned a group of people and projects that had and continue to have a long term effect. ROS provides the backbone for most modern robotics research and you're probably reading this on something that can trace its intellectual lineage back to PARC. On the other side of the coin, there is also a class of organizations that have lasted for decades. You could easily weave them into a success narrative, but they somehow smell 'off', like something that should have been tossed out long ago. And of course there's everything in between.

Specific Failure Scenarios

The two clear-cut failure scenarios are a failure to launch and dying before doing anything useful. As discussed above, a more insidious scenario is becoming a 'zombie' organization — consuming money and mindshare (and perhaps even thinking we're a live player) while ultimately going nowhere.

Below I will dig into the reasons *why* those things might happen. Teasing out how one would know that one has fallen into a particular failure mode is more important than enumerating failure modes is. I will take a stab at answering "how will we know how we failed?" but you should consider it an open question.

Failure modes fall into two buckets — one set of failures apply to DARPA-riffs in

²⁷⁵ While Xerox PARC technically still exists, after a leadership change in 1983 it arguably morphed into a different organization.

general and the other set applies specifically to PARPA. It's important distinguish between these two categories because I don't want PARPA's failure to condemn the idea of riffing on DARPA as a whole, but at the same time it's foolish to cling to a failed theory.

Why would riffing on DARPA be a bad idea?

Structure doesn't actually matter

Institutional structure could be much less important to an organization's ability to enable new things than I've argued for. Instead of structure, the factors that determine the activities an organization can do well could depend almost exclusively on its particulars — its leaders, its funders, its employees, its time and place. It could also be that structure does matter but that focusing on DARPA is zooming in on the entirely wrong place on the map.

The current ecosystem has no unfilled niches

One of the core premises behind riffing on DARPA is the assertion that overlapping institutional constraints rule out several classes of creative work and therefore we need new institutional structures to fill the role that industrial labs once occupied. This could be wrong! It could be that enabling most physically-possible technology just takes more people Elon Musk threading needles with the institutional structures we already have. More broadly this would mean that existing institutional structures have fewer structural constraints than it seems like they do.

A DARPA-riff is unable to institutionally decouple from papers or profit

A DARPA-riff might be inevitably coupled to other institutions in a way that prevents it from having significantly different constraints. Any DARPA-riff will inevitably need to work with academia for early-stage research and profit-focused companies to diffuse technology. It could be impossible to both work with these other organization and stand apart from their incentives, slowly but inexorably being pulled to publish or profit.

DARPA-riffs need government funding

The hypothesis that 21st century riffs on DARPA should be private could be wrong and it's impossible for a DARPA-riff to succeed without government funding. Perhaps the Department of Defense (and maybe the Department of Homeland security) are the only places where massive budgets and are sufficiently aligned with outputs that don't necessarily generate papers or profits. One could argue that governments are the only institutions that have the timescale (the US government is a relatively young 245) and scope (GDP) to capture the value from long-term, high-uncertainty programs.

Coordinated Programs are Bunk

It could be that DARPA-style programs are a holdover from the Cold War that don't actually help innovations exist. In our world of shared documents, agile development, and cloud labs the payoffs for research management could have decreased significantly. Instead of trying to direct research work towards a goal, the correct strategy in all domains might just be "give good people money and let them rip!"

Why would PARPA in particular be unsuccessful?

Size

We could fail because a small new institution wouldn't be able to move the needle. We're working off of the hypothesis that it is important for big things to start small, but perhaps a successful DARPA-riff needs to start big. 'Size' could be a failure mode either because to succeed we actually need to start big or that the organization fails to grow (for one of a number of reasons) and fails to hit the critical mass of programs necessary for a DARPA-like portfolio.

Working on the wrong things

PARPA could fail to work on the 'right' programs. A causal factor (probably people) could make us consistently bad at designing or picking programs. We could also fall into a trap of working on things that are 'easy' instead of 'right' —

an easy program might fit nicely into a budget, be part of a hype-wave, or just require convincing fewer people.

This trap is insidious because research has a halting problem — you can never be sure that a project won't produce something amazing after dragging on for a long time. This goes for the organization as a whole. However, at some point, even if you have faith in the general DARPA-riff model, it will make sense to kill PARPA after too many failures. It may be impossible to know if it's sheer bad luck or something systemic.

Path dependence

Working on the right things in the wrong order could also lead to failure. If there is fundamental uncertainty about how a program will turn out, you can end up getting a whole bunch of duds in a row. HHTTTT and TTTTHH have the same failure rate, but a string of early successes could buy enough trust that people could roll with a string of failures while a string of early failures could kill the organization before it had a chance to get to the successes.

People

It could just be that I am the wrong person to do this. I bring neither experience managing coordinated research programs nor capital to the table. I could be insufficiently charismatic or outgoing to convince the right people to buy into the project before it has demonstrated success and has its own momentum. I can try to mitigate this by working with other people, but *those* people need to be convinced as well. It could be that there is a group of people with whom the idea would work but they end up not being available at the right time. Perhaps they will read this in the future and succeed where I fail. That would be a good outcome.

Shifting Technology from Impossible to Inevitable

We end our exploration of PARPA's idea maze at last.

The maze exists in the first place because constraints on existing institutions have created a gap in the innovation ecosystem. 'Solutions R&D' was once the purview of industrial labs, but no longer. Shifts in both technology and how companies work mean that we can't expect Bell Labs' successor to look anything like Bell Labs. Nor can we count on the two institutions we've defaulted to in the 21st century. Solutions R&D's tense mix of pragmatic application-focus and piddling around is too researchy for startups and too engineering heavy for academia. So, with a respectful hat-tip to Chesterton's fence, we need something new.

It's impossible to know the true layout of the idea maze before traversing it, but we can make many hypotheses about its twists and turns. While we could start from absolute scratch, it's worthwhile to pay attention to the solutions R&D heroes of yore — we can't follow in Bell Labs' footsteps but it might be possible to follow in DARPA's. There are many junctures that anybody building a DARPA-riff will need to consider: experiments, program managers, timescales, sales channels, missions and more. Perhaps the thorniest question we'll all need to face is "How does money work?" Unfortunately, if you draw a box around all the activities in impactful solutions R&D, it's probably not a profitable venture (value creation and value capture are, alas, not always the same). However, there are plenty of strategies that a DARPA-riff can use to get enough money to continue through the maze!

Finally, I want to propose a specific path through the maze — PARPA. We can look at PARPA²⁷⁶'s potential path through the idea maze as a sequence of hypotheses — junctures that decrease your risk of failing for each one you successfully pass. Bundled together, PARPA's path through the maze looks like:

²⁷⁶ The P is for private, but you can also imagine that it stands for Phenomenal or Prodigious if you want.

1. Create and stress-test unintuitive research programs in a systematic (and therefore repeatable) way.
2. Use that creditability to run a handful of research programs and produce results that wouldn't happen otherwise.
3. Use that creditability to run more research programs and help them 'graduate' to effective next steps.
4. Make the entire cycle eventually-self-catalyzing by plowing windfalls into an endowment.

Each of the first three steps roughly corresponds to a sequence of organizational 'forms' that PARPA will evolve along as it validates hypotheses (or invalidates them and has to figure out new ones).

Any idea maze is fraught with danger — many who enter do not survive and many who survive make it through by the skin of their teeth without real success. There are many ways that PARPA can both succeed and fail; it's important to acknowledge this before setting out so that the journey is not just about its outcome but can help other people traverse the same or similar mazes.

I hope to have convinced you of several points:

- It's possible to create technology that expands human capability and wonder.
- Creating this technology requires (among other things!) a set of activities that form a 'solutions R&D' niche in the innovation ecosystem.
- Built-in constraints prevent modern corporate R&D, academia, and startups from doing a good job at filling the solutions R&D niche.
- A significant chunk of these constraints has to do with the fact that while profit is important, value capture can hamstring a lot of solutions R&D work.
- We need to try new organizational and legal models to do good work that is "too researchy for startups and too engineering-heavy for startups."

- DARPA is a good starting-point for creating new organizational and legal models to fill the solutions R&D niche.
- 21st century DARPA-riffs need to exist outside of 21st century government bureaucracies.
- There are many ways to build a DARPA-riff.
- There is a set of things every DARPA-riff will need to consider, including PMs, timescales, sales channels, missions, and perhaps most importantly — money.
- There are many ways that a DARPA-riff can make money, each creating a set of incentives and constraints.
- Overall a DARPA-riff that is doing unique work will probably not be profitable though a portion of it might be.
- Part of a DARPA-riff's role is to act as a serious context-of-use for experiments in research management, for example “it might be possible to develop a discipline around program design.”
- A legal structure comprising a (big-N) Nonprofit and a C-corp will give PARPA the best chance of success and avoiding incentive-traps.

If you think these are wrong, show it! Do great solutions R&D within existing institutions! Build a successor to Bell Labs! I like being right, but I would like those things more.

If however, you think the ideas I've presented are plausible, let's put them into action. There are so many things to be done. You can put these ideas into practice at an existing institution or use them to build a new institution — either a DARPA-riff or something entirely different. There are also many concrete ways to help PARPA! Reach out if you or someone you know would make an excellent PM or otherwise want to put legwork into building the organization. Send precise hunches about projects and people who could form the seed of a program. Donate or invest. Or just send this piece to someone who will appreciate it.

Together we can *shift technology from impossible to inevitable*.

Gratitude

Ideas don't spring out of a single brain, and this piece is no different.

Many excellent people read and helped refine early drafts: Adam Marblestone, Jed McCaleb, Pamela Vagata, Andy Matuschak, Mark McGranaghan, José Luis Ricón, Cheryl Reinhardt, Sam Arbesman, Michael Nielsen, Nathan Ihara, Martin Permin, Luke Constable, Arnaud Schenk.

We stand on the shoulders of too many giants to give them all credit, but to list a few who helped me grope through the idea maze: Marissa Weichman, Ilan Gur, Semon Rezchikov, Tim Hwang, Jeff Graham, Arathi Prabyankar, Patrick Collison, Malcolm Handley, Olivia Wang, Noah Tye, Rebecca Li, Steven Glinert, Keegan McNamara, Michael Filler, Matt Clifford, Adrienne Little, Peter Van Hardenberg, Josh Tobin, Cameron Kelly, Evan Myazono, Jeff Lipton, Lee Ricketson, Adit Swarup, Alexey Guzey, Sebastian Winther, Luke Durant, Rachel Zucker.

The styling and interface of this website is mostly derived from Andy Matuschak and Michael Nielsen, "How can we develop transformative tools for thought?", <https://numinous.productions/tft>, San Francisco (2019)

This work was graciously sponsored by [The Astera Institute](#).

If you want to support PARPA with either your time or money, [please reach out!](#)