Harvard Business School

624-065
DECEMBER 11, 2023

TEACHING NOTE

X: The Foghorn Decision

Case Synopsis

In February 2016, Kathy Hannun—a project leader at X, Alphabet Inc.'s "moonshot factory"—had to prepare a recommendation for senior leadership regarding the future of Foghorn, a project she was leading to develop a carbon-neutral process for converting sea water into fuel. Recognizing the blueprint for projects at X required, (1) addressing a huge problem with a (2) radical solution using (3) breakthrough technology, Hannun had to decide whether to recommend killing the project or continuing to push forward.

This case examines how culture and incentives can be used to guide innovators' efforts, debates the value of "learning curves," emphasizes the importance of understanding opportunity cost, and highlights the challenges of evaluating innovative processes.

Case Placement

This case can be used both as an introduction to the challenges of managing innovation in general as well as an example of strategies that large organizations can use to develop radical innovations. At the Harvard Business School, this case has been taught in the first-year Technology and Operations Management course as a part of the innovation module as well as in the second-year Tough Tech Ventures course. The case has been taught to executives as a way of discussing organizations designed for radical innovation and to graduate student scientists and engineers (e.g., PhDs and post-docs) in the physical and natural sciences as a way of introducing managerial frameworks for their own work.

Learning Objectives

The Foghorn case is designed to help students understand the following concepts:

- Compared to standard processes, where managers often seek to reduce the variance of output, radical innovation processes often focus on increasing the variance of output to increase the odds of an extreme value outcome.

This note was prepared by HBS Professor Kyle Myers and independent researcher Walter Frick (Nonrival) for the sole purpose of aiding classroom instructors in the use of "[Case/Course Title for Disclaimer/Footer]," HBS No. [Case Number]. Funding for the development of this note was provided by Harvard Business School and not by the company. It provides analysis and questions that are intended to present alternative approaches to deepening students' comprehension of business issues and energizing classroom discussion. HBS cases are developed solely as the basis for class discussion. Cases are not intended to serve as endorsements, sources of primary data, or illustrations of effective or ineffective management.

- When a process involves inputs and outputs that are unobservable, rare, and uncertain, as in the case of radical innovation efforts, managers must rely on theoretical frameworks for design and evaluation.
- The sequencing of tasks in innovative projects is closely linked to the size of the opportunity costs the manager is facing and the spillovers those tasks might generate. In the context of radical innovation, these opportunity costs and spillovers can be very large.
- Designing processes that involve frequent iteration across projects (i.e., "failing fast") requires a unique blend of employee selection, culture, and incentives that can counterbalance the default, intuitive ways that many people approach and evaluate their work.
- Forecasting a project's progress based on prior work with a "learning curve" can be more accurate when the curve shows a causal effect of some input, which the manager controls, on the focal measure of progress.

Supplemental Materials & Reading

Supplemental Material

- A PowerPoint slide version of **TN Figure 3** is included for use in the classroom during the learning curve discussion. Note: the graph and discussion of learning curves are not included in the case study. The instructor can show the slide or draw the graph on the board.
- An Excel document containing a highly stylized net present value analysis of the Foghorn project at the time of the case.

Supplemental Readings

- Edmondson, A. C. (2023) *Right Kind of Wrong: The Science of Failing Well*. Simon and Schuster: New York City.
- Gertner, J. (2012). *The Idea Factory: Bell Labs and the Great Age of American Innovation*. Penguin: London.
- Hill, L. A., Brandeau, G., Truelove, E., Lineback, K. (2014). *Collective Genius: The Art and Practice of Leading Innovation*. Harvard Business Review Press: Boston.
- Cowen, T., Gross, D. (2022). *Talent: How to Identify Energizers, Creatives, and Winners Around the World*. Macmillan: New York City.

Suggested Assignment Questions

1. Should Kathy Hannun kill the Foghorn project? If so, how exactly should the project be cancelled? If not, what rules should determine how the project progresses?
2. How has X designed its operations to facilitate rapid iteration through ideas? What sorts of ideas might these processes lead X to overlook?
3. How should X evaluate its employees? How should Alphabet evaluate its investments in X?

Case Analysis

The analysis of this case is presented in five sections. The first two sections are particularly useful if the instructor is using the Foghorn case as way of introducing the differences between more "standard" and more "innovative" production processes. Section 1 clarifies the distinction between standard and innovative processes, and it illustrates the connection between the manager's objectives and how they should conceptualize their process flows. Section 2 highlights the challenge of evaluating innovative processes, which motivates the importance of theoretical models when managing innovation.

The next three sections highlight specific theoretical frameworks that are relevant to this case. Section 3 discussing the sequencing of operations in innovation processes (e.g., which components of a project should be focused on first?). Section 4 discusses the role of a culture and the design of incentives that promote "failure". Finally, Section 5 examines what attributes of learning curves make them helpful for allocating resources and forecasting future progress.

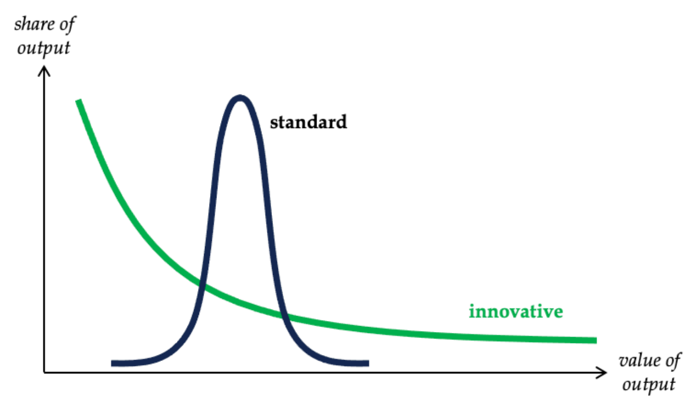
Section 1: Standard versus Innovative Production Processes

It is useful to take an operations management view of production processes to set the stage for the Foghorn discussion. Doing so motivates a contrast of *standard* production processes and *innovative* production processes. In what follows, it may be helpful to conceptualize these processes as the workflow of a single team or project.

A standard production process is typified by the use of low-uncertainty methods to create multiple copies of an existing product. In more standard processes, the inputs and outputs have clear specifications, and all of the output (i.e., the products) are intended to be sold to customers. Thus, the manager's objective is to *maximize the sum* of the value of all output (e.g., the price each product can be sold at) less the total costs of production.

Conversely, in an innovative production process, the objective is to generate new ideas that could form the basis for new products (that can then be produced in a standard process). But not all of these new ideas are intended to be sold, just the few that will form the basis of the new product. This changes the optimal design of the process considerably. Consider a simple case where one new idea is needed for one new product. In this stylized example, the manager's objective is to *maximize the maximum* value of all output (e.g., the expected value of the single best idea) less the total costs of production.

TN Figure 1 Standard and Innovative Process Output Distributions



Source: Case author.

A useful way to visualize this difference between standard and innovative processes is to compare the distributions of the value that is produced— "value" here could reflect be the price-cost margin of a single product produced by a standard process, or the net present value of developing a new product line based on the ideas produced by an innovative process (see **TN Exhibit 1**). In standard processes, a goal is often to *minimize or cap the variance* of the output produced to ensure that all items are within spec and can be sold. A manufacturer that produces consumer goods typically wants to produce output that, given some specifications, are as close to identical as possible in order to maximize the sum of the value of all goods.

Conversely, in innovation processes, only a few new ideas may be necessary, perhaps just one. Now, the manager's goal is to *increase the variance* of the output produced to ensure that "the best" idea is identified. Increasing the variance of output, especially in the context of idea generation, often comes at the cost of producing a lot of low-quality output. But paying this cost may be worthwhile if it increases the odds of an extreme value outcome (e.g., by learning about the faults of one idea, a team may be able to produce more valuable ideas in the future).

This view of standard and innovation processes (**TN Figures 1**) is a useful primer for the X case because it formalizes the motivation for many of the otherwise peculiar systems that X, and other innovative organizations, had in place: these organizations were trying to increase variance and maximize the maximum output of their processes.

Section 2: Evaluation

What do the differences between standard and innovative processes imply for the evaluation of these processes? Innovative processes center around "ideas," and ideas are difficult to characterize: ideas themselves are hard to measure; the value of ideas is often very uncertain; even when an idea is well-defined, it is difficult to know how much effort has been invested in it (e.g., how does one measure time spent *thinking*?).

It is useful to consider the degree to which the inputs and outputs of these processes are observable and involve risks (see **TN Figure 2**).[a] Standard processes are characterized by a high degree of observability and a low degree of risk. In the limit, if the process is perfectly observable and there is no risk, then evaluating the process's performance or understanding the effect of a change to the processes can be calculated using *mathematical or statistical models and simulations* and directly seeing how the output of the process changes.

As managers start to lose their ability to clearly see the process or uncertainties increase, they must rely on alternative approaches. Often this involves the use of *experiments* that manipulate the inputs of a process and then observing changes in *proxies* for the true outcome of interest.

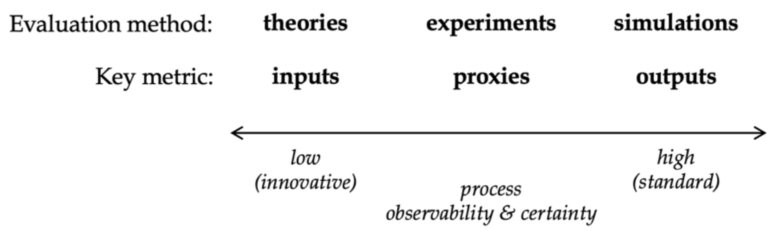
In the pursuit of innovation, the observability and certainty of a process tends to be very low. As simulation and experimentation becomes less feasible, managers of innovative processes must fall back onto their *theories* and use what they can see about the *inputs* of their processes to evaluate themselves. What is a "theory" here? It is a hypothesis that if the process is designed a certain way and certain inputs are provided, then it will generate a certain output in expectation.[b] As Teller often put it in his

[a] "Risk" here summarily refers to both known uncertainties (i.e., objective probabilistic outcomes) and unknown uncertainties (i.e., subjective ambiguities).

[b] Note, this output may still be probabilistic. For instance, if the total value to Alphabet generated by X never exceeded their total costs, that would not necessarily imply that investment in X was inefficient. Rather, it may imply that X was a risky gamble that just happened, ex-post, not to pay off.

public remarks about how he managed X, his focus was on "habits, not outcomes." He judged X based on what he could observe about their inputs and his theory of how those inputs would generate innovations.

TN Figure 2 Evaluating Innovation versus Standard Processes



Source: Case author.

Clearly, Teller focused a large amount of his efforts on maintaining his own theory of "systematizing innovation," as he often referred to it. As evidence by his "Operating Manual"[1] and "Tips for Unleashing Radical Creativity"[2], his theory involved the recruitment of "creative and weird souls" as employees and it did not involve the extensive tracking of inputs and outputs—there were no "OKR (Objectives and Key Result" metrics such as those that permeated Google's main business lines.

Rather, Teller's approach was full of hypotheses. For example, X frequently highlighted the extreme diversity of their workforce, which included experts ranging from rocket scientists to marine biologists, to puppeteers, to concert pianists. There is good evidence that more diverse R&D teams are more productive,[3] but X was operating well beyond the scope of normal workforce diversity driven by Teller's theory. Another theory touted by Teller was the value of encouraging his workforce to think only of large-scale problems and solutions. As he once put it, "it's often easier to make something 10 times better than it is to make it 10 percent better."[4] Again, this was not a claim rooted in any formal quantitative analyses. Rather, it was a theory about the ways in which innovators' objectives influence the types of ideas they generate, or even the types of innovators who wanted to work at X in the first place.

The broader lesson here is that, as a process becomes more unobservable and more uncertain, the manager will eventually need to rely on theories when designing and evaluating the process. Furthermore, these managers should then be judged more on the validity of these theories and less on whether the desired output was actually produced. What can inform these theories of innovation? The analysis below provides some specific theoretical frameworks related to the Foghorn case. Beyond this case, there is a growing body of academic work that formalizes many of the tradeoffs of managing innovators,[5] and there are many useful histories and case studies of other organizations that have tested their own theories of innovation.[6] The instructor can incorporate some of those research findings into their teaching plan if desired.

If this case is being taught in a course that has already included multiple discussions surrounding innovative processes, the material in this and Section 1 may not need to be included in the case discussion as suggested below in the Teaching Plan. But for courses where this case discussion will be one of the first about managing innovation, this comparison to standard processes can provide a useful transition. In both scenarios, the frameworks motivate the need for unique managerial practices in the world of innovation.

Section 3: Sequencing Operations

A key question for teams engaging with a new R&D project is which component of the project to focus on first. In the context of Foghorn, there were an enormous number of components that the team could have engaged with (e.g., sourcing seawater, sourcing hydrogen, creating durable membranes, connecting to the global fuel supply chain, etc.). How should the Foghorn team have chosen what to prioritize?

The academic literature on this question has yielded a resounding "it depends",[7] but a focal point of this case is to emphasize, as described below, the contrasting roles of opportunity costs and learning via spillovers. To highlight these two forces, it is useful to classify components of an R&D project along three dimensions, as follows:

(1) *feasibility* (i.e., the probability that the component can successfully be produced as desired),

(2) *direct impact* (i.e., the difference in the value of the focal product if the component has the desired attributes), and

(3) *foundational value* (i.e., the value of the knowledge learned while working on the component).[c]

The question is: should managers prioritize components that are more or less feasible, have higher or lower direct impact, and higher or lower foundational value? Intuition might suggest prioritizing components that are high on all dimensions—more feasible, larger direct impact, and larger foundational value.

First, let's consider feasibility and direct impact. X's mantra of "tackling the monkey first" reflected a strategy of pursuing components with low feasibility and high direct impact.[d] This was driven by a strategy to minimize opportunity costs. By focusing on low feasibility and high direct impact components, fewer resources could be sunk into a project before learning more about the project's true value.[e] And X's opportunity costs were large. They had a relatively small workforce with an extensive set of generalizable skills. As Rich DeVaul, X's Director of Mad Science, put it in the case, X employees "are people who know less and less about more and more." Thus, by components with low feasibility and high direct impact, this small, flexible workforce could avoid paying large opportunity costs.

The third key dimension is the notion of foundational value, which need not be correlated with feasibility or direct impact. There may be project components that, simply by engaging with them, could lead to indirect value through "knowledge spillovers". For example, new ideas for how to approach the focal project or an entirely new project may result. Such spillovers from R&D activities have even been shown to flow far outside the boundaries of innovative firms and to be quite large.[8]

In the case, the story of Project Loon provides an example of the surprising foundational value of some components. The data generated by millions of miles of test flights ultimately proved a useful input into algorithms that could help steer balloons in the stratosphere. Notably, the data from early test flights was not prioritized as a project component with foundational value, but it ultimately proved

---

[c] An important note is that each of these dimensions are being considered here conditional on some amount of investment (e.g., feasibility given a financial investment of some dollar amount).

[d] Conversely, in Teller's analogy about training a monkey to talk while standing on a pedestal, the pedestal is the high feasibility low direct impact component.

[e] An intuitive, albeit imperfect, analogy can be made here to the focus on identifying bottlenecks in standard process analyses. In a standard process analysis, resources are directed towards the bottleneck step of a system because it sets the capacity of the entire system. Here, resources are being directed to the components of a project that have the largest effect on the expected value of the project; loosely speaking, low feasibility high direct impact components are the bottleneck.

very useful in moving the project forward. This example highlights the notorious difficulty of predicting the foundational value of any R&D project component and it sets the stage for a discussion in the class about what sorts of ideas X might have missed because of its choice to focus primarily on considerations of feasibility and direct impact and less on considerations of foundational value.

Section 4: A Culture of, and Incentives for, Failure

A primary challenge for managers of innovative processes is influencing the direction of the creative workforces they oversee. X's scientists and engineers possessed unqiue skills and knowledge often far beyond that of the manager, suggesting there would be gains to giving project control to the workforce. But these scientists and engineers also may have preferences over the nature of their work that were not well aligned with organizational priorities.

Because of innovators' specialization and unique preferences, as well as biases like the sunk-cost fallacy, the reluctance to move from one project to the next can be very difficult to overcome.[9] At the same time, managers often possess superior information about signals of market demand or the broader capabilities of the firm and its strategic plans. Thus, managers faces a tradeoff of relinquishing control to their scientists and engineers in order to leverage their highly specialized knowledge and skills versus minimizing the risk that their workforce pursues ideas less valuable to the organization.[10]

The leadership of X had chosen to balance this tradeoff with several "cultural" tools centered around the idea of *killing projects as fast as possible*. That is not to say X promoted failures writ large. Rather, an enormous focus was placed on, what Edmondson (2023) refers to as, "intelligent failures" that were the result of thoughtful experiments designed to learn the true value of a new idea.[11]

There was a clear set of high-powered incentives (e.g., bonuses for killing projects) as well as a lack of some incentives one might expect (e.g., employees have no explicit or implicit equity in the success of very early-stage projects). The X campus was physically separated from the main campus of Google, which mirrored a strategy taken by prior "skunkworks" such as Bell Labs and PARC. The recruitment of new X employees emphasized identifying individuals who exhibited a high tolerance for ambiguity and risk, as well as demonstrating a flexible approach to their work.

Perhaps most noticeably, Teller had spent considerable effort to develop a culture that celebrated failure. For example, there were celebratory parties when projects were cancelled. As Teller noted: "A lot of this was about the attitudes, social norms, and a commitment to critically thinking and finding ways to be passionate and dispassionate in the same moments, or at least in the same month."[f] Implicitly, Teller's "passionate-dispassionate" oxymoron expressed his desire for X employees to be committed and passionate about the broader mission of X while also being willing to dispassionately end their work on any specific project if targets were not met. In this way, Teller recognized that a project failing did not imply that any suboptimal decisions were made. Furthermore, he recognized that capturing the learnings from these failures would require a commitment to the broader mission of X.[12]

The mission statement of X, with its three criteria, is a particularly interesting tool to discuss. One way to view the role of X's mission statement is as a tool for shaping and rewarding X's employees as they decided which projects to pursue.[13] By specifying the objectives of X in the way that they did, managers were attempting to both shape employee's preferences and provide non-monetary compensation for their effort when it was exerted in the direction of the mission. Thus, the mission statement provided a guide for employees as to how to allocate their efforts across many possible tasks.

[f] Teller often refers to himself as a "cultural entrepreneur" in public engagements, emphasizing the importance he places on this topic (e.g., see: https://www.youtube.com/watch?v=adLiT8JAc3E).

For instance, if an employee was nearly indifferent as to which of two projects to pursue, but could only pursue one, the mission statement suggested which of those two should be pursued (the one that more closely aligns with the mission).

As evidence of the role of organizations' missions when attracting talent, the following excerpt from a news article covering competition between OpenAI and Google for talent includes a quote of Ilya Sutskever, who left Google and became the Chief Scientist at OpenAI:

> *"They did make it very compelling for me to stay, so it wasn't an easy decision," Sutskever says of Google, his former employer. "But in the end, I decided to go with OpenAI, partly of because of the very strong group of people and, to a very large extent, because of its mission."*[14]

Overall, the mission statement encouraged more effort in a particular direction that X's leadership believed to be most valuable for the Alphabet enterprise. In this way, it also helped Alphabet ensure that any foundational value that was generated by X's projects was more likely to be captured, which helps address the challenge of appropriating value from the spillovers of these projects as discussed in Section 3 above. The case includes a quote from Astro discussing how some lessons learned and ideas born out of prior projects were put "in the refrigerator," because they weren't immediately useful but showed some promise of having foundational value.

However, it is unclear whether the specific mission X chose might have led them to pass on otherwise valuable ideas. This question can be used to spark in-class debate, as highlighted in teaching plan below.

Section 5: Causal Learning Curves and Forecasting Progress

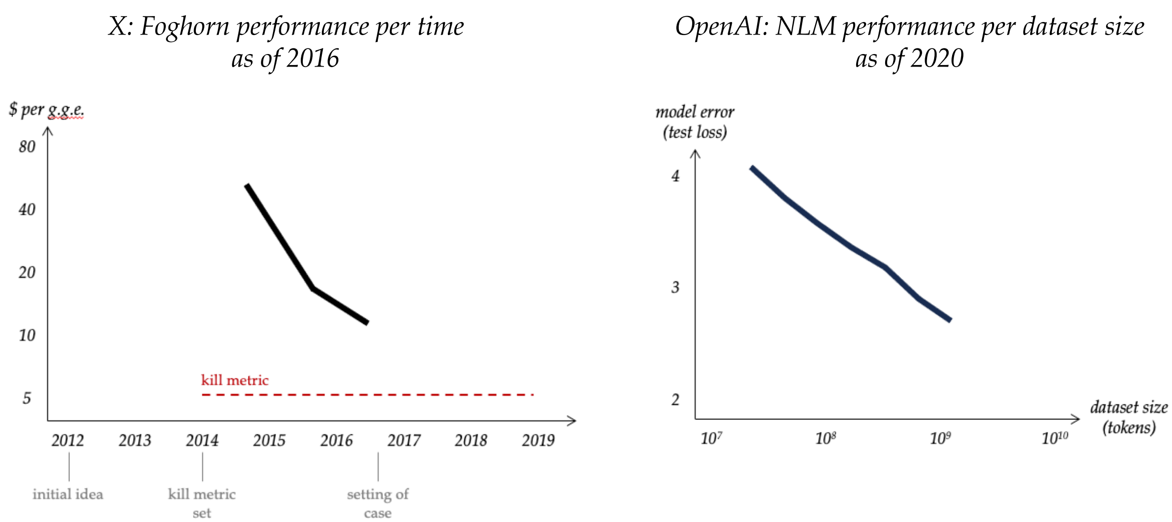
Learning curves are commonly used to illustrate the progress of ongoing R&D projects. They plot some measure of progress (e.g., efficiency) against some measure of input (e.g., time, investment). **TN Figure 3** uses Exhibit 10 of the case to construct a learning curve for the Foghorn project; it plots Foghorn's efficiency, in terms of dollars per gas-gallon-equivalent (lower is better), as a function of time. For comparison, **TN Figure 3** also illustrates a learning curve for Neural Language Models (NLMs; also referred to as Large Language Models or "LLMs") at the company OpenAI, which has been adapted from a 2020 academic publication co-authored by OpenAI employees; it plots the accuracy of the NLM, in terms of test loss (lower is better), as a function of the size of the data used to train the model.

It is intuitively attractive to use learning curves such as those in **TN Figure 3** to forecast future progress (along the dimension of the *x*-axis). But it is crucial to note that such forecasting is much more useful when the relationship is a *causal learning curve*; that is, a learning curve for which the independent variable (the *x*-axis) represents an input into the innovative production process that the manager has (ideally, experimentally) manipulated and has some understanding of how that increasing or decreasing that input can *cause* progress rather than being just correlated with it.

Many learning curves, as in the case of the Foghorn curve, are drawn as a function of time. However, such a curve provides little information as to exactly how progress is being achieved over time—it *does not* illustrate a clear causal relationship between a specific component of the Foghorn project and the efficiency of the technology. Extrapolation of the curve *suggests* progress may continue, but it provides no information to the manager about how specifically to allocate resources to *cause* additional progress.

Now for comparison, consider the OpenAI NLM learning curve. The *x*-axis is the size of the dataset used to train the NLM and it shows an approximately log-linear relationship between the size of the data and the model's error. This provides OpenAI with a clear view of a component—dataset size—that appears to *cause* progress. OpenAI's managers, scientists, and engineers may not perfectly understand this relationship, but they can clearly theorize about it and continue to run experiments on the dataset size component of their project. Conversely, the Foghorn learning curve by itself does not clearly motivate any new experiments. It is perhaps not surprising then that OpenAI was willing to dramatically increase its investments in scaling the size of their training datasets whereas X ultimately chose to kill Foghorn. OpenAI was working with a more causal learning curve than X.[g]

TN Figure 3 Learning Curve Comparison



Source: Case author, including adaptation of Figure 1 of Kaplan et al. (2020). Scaling laws for neural language models. Accessed at https://arxiv.org/pdf/2001.08361.pdf.

Note: NLM stands for Neural Language Model. The scales of the *y*-axis in the Foghorn plot and the *x*-axis in the OpenAI plot are logarithmic. In both plots, smaller values on the *y*-axis indicate improved efficiency of the technology.

[g] It is difficult to pin down the specific timeline of OpenAI's investments in its NLMs, but the data used to generate the plot in **TN Figure 3** roughly corresponds to the timing of the GPT-2 to GPT-3 transition. Detailed breakdowns of OpenAI's investments are not available (i.e., separating the cost of data acquisition versus training runs), but the best estimates suggest it cost approximately $50,000 to train GPT-2 and approximately $5 million to train GPT-3.g These magnitudes are roughly on the same order of magnitude of what X might have considered when debating X in the case.

Teaching Plan

An 80-minute class session can be organized into the following parts described in detail below:

Part	Length
1. Introduction	5 minutes
2. Ordering operations & opportunity costs	10 minutes
3. X's operating model: Incentives & culture	10 minutes
4. Mission statement and missing ideas	15 minutes
5. Evaluating innovative processes	15 minutes
6. Foghorn debate & learning curves	20 minutes
7. Summary	5 minutes

1. *Introduction (5 minutes)*

To open the class, the instructor can first provide roadmap of the discussion to the students to set their expectations, noting that there will be short debate about the Foghorn project at the start of the class, then a considerable amount of time will be spent focusing on X as a whole, before again debating the specifics of the Foghorn project.

To provide some context, the instructor can emphasize the point that X is but one of numerous "skunkworks"—isolated, R&D laboratories with an objective of delivering innovation—that exist at the time of the case across a wide range of industries (e.g., Amazon Lab 126, Ford Special Vehicle Team, Walmart Labs, Nordstrom Innovation Lab, Facebook Reality Labs, Nike Sport Research Lab, etc.). Furthermore, as noted in the case, forerunners can be seen throughout the 20th century in Xerox's PARC and Bell Labs, and even looking back to the laboratories of Tesla and Edison.

Focusing briefly on PARC and Bell Labs provides a useful way to set tension for the discussion about Foghorn and X. In the cases, these corporate laboratories struggled to sustain innovation. Bell Labs was responsible for some of the most important inventions of the 20th century: the transistor, the laser, satellites, photovoltaic cells, and cellular networks. PARC was responsible for some of key inventions surrounding laser printers, ethernet, computer graphics, and the personal computer (i.e., a window-based, mouse operated user interface). However, the antitrust split of AT&T in the 1980s clearly altered Bell Labs' strategies and operations in a way that yielded fewer innovations of the same scale as in decades prior, and PARC was unable to capture much (or sometimes any) of the value of their biggest ideas. Bell Labs and PARC both still existed at the time of the case, but in both cases they had faded from their prominence. This motivates the question of whether X will suffer the same fate or not.

To center the discussion squarely on X and Foghorn, the instructor can ask for two students to provide alternative views as to whether Kathy Hannun should kill Foghorn or not, a sort of opening arguments to set the stage. It is helpful to note the reasonings put forward by these two students because you can return to them throughout the class to see if their beliefs change as the class progresses.

> **Q:** *Who can provide a clear, comprehensive "opening argument" as to why Foghorn should be killed? And who can do the opposite as to why X should keep Foghorn alive?*

After this, the instructor can take a class-wide poll as to whether they think Foghorn should be killed or not. Quite often, this ratio is near 50:50, making the uncertainty of this decision clear to all. One option to ensure variation in responses to the poll is to offer four options: Strong-continue, Weak-continue, Weak-kill, Strong-kill.

As a last bit of introduction, the instructor can ask the following:

> **Q:** *Why can't X just calculate the Net Present Value (NPV) of Foghorn and use that? Surely X's parent, Alphabet, is not capital constrained; they have an estimated $75 billion in cash on hand around the time of case.*[15] *So, isn't this just a question of NPV?*

This question often yields answers related to the large uncertainties underlying NPV estimates of projects like Foghorn. Yes, one could construct an NPV estimate, but the confidence intervals of that estimate may reach far into negative and positive values. To fully reinforce this point, the instructor can use the supplemental Excel document that contains a highly stylized NPV analysis of Foghorn based on a range of assumptions that can be modified. Using this document, one can obtain seemingly plausible NPV estimates that span from a complete waste to generating impressive returns. This provides a useful bridge into the next part.

2. *Ordering operations & opportunity costs (10 minutes)*

Building off of the discussion of the limitations of NPV estimates, the instructor can open this part with the following:

> *Instead of relying on NPV estimates, the case references X's use of "kill metrics" and "milestones." I'm sure you've seen or been a part of projects with milestones where, for example, you receive a steady flow of resources, and if you ever complete your first goal, then you get more resources towards the pursuit of the next goal. But I'm guessing many of you haven't seen or been a part of projects with kill metrics where, if you don't complete your first goal in a specified amount of time, you are committed to shutting the project down.*
>
> **Q:** *Does anyone have any experience with something that looked like a kill metric in the projects they've been a part of?*

Typically, only a few students have ever encountered incentives that resemble kill metrics. Which itself is a point the instructor can emphasize about how non-intuitive such incentives are. Next, the instructor can drive the class towards a discussion about how to order the tasks within an R&D project with the following:

> *I'd like to talk about how these kill metrics help focus a team's effort on a subset of the many possible features of a new project. Clearly, with Foghorn's kill metric of 5 $/g.g.e. within 5 years, Hannun's team wasn't worried about certain challenges like how they'd integrate their technology into the fuel supply chain, or the physical size or layout of the system, etc. The kill metric focused the team's efforts on a specific measure of efficiency. But what exactly should the work on first?*
>
> **Q:** *More generally, let's imagine you're in charge of a new project at X and are devising your own kill metric. What features or components of the project would you want your team to focus on first?*

Here, the instructor can either force the class to immediately consider the dimensions of feasibility, direct impact, and foundational value by iteratively defining the dimension and then asking if X should prioritize components that are high or low on the dimension; or, the instructor can solicit these ideas from the students. In each case, the instructor should ask ***why*** they think components that are high or low on a dimension should be prioritized. Usually, in the case of feasibility and direct impact, students' first responses will reflect opportunity costs to a certain degree, but it can be helpful to push for clarifications until it is made explicit. This exercise is captured as shown in *Board #1* (**TN Appendix A**).

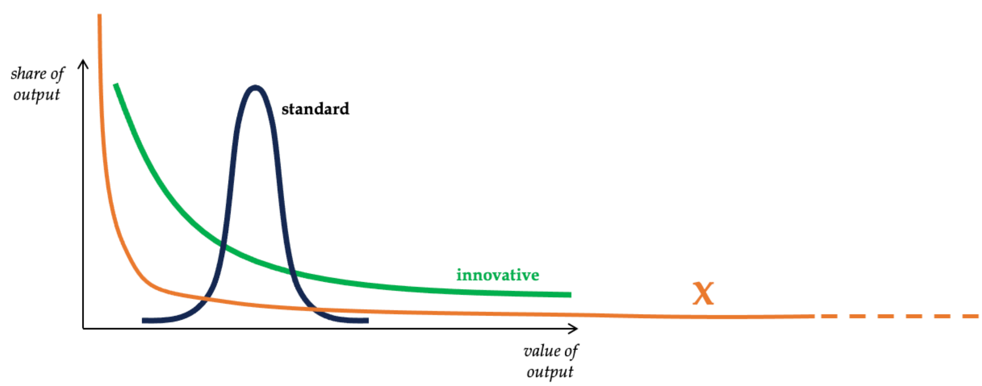
The notion of spillovers from foundationally valuable components being driven by the cumulative nature of research is typically apparent to students quite quickly. But if time permits, one way to emphasize this point further is to note the power of combinatorics—for example, let's say a "product" is a set of one or more "ideas"; if we have 2 ideas, we can make 3 products; with 3 ideas, we can make 7 products; with 7 ideas, we can make 127 products, and so on.[16]

After completing the "Why?" component of *Board #1*, the instructor can push on the idea of opportunity costs and ask:

Q: *What were the determinants of X's opportunity costs?*

The goal with this question is to uncover the importance of how many resources an organization has (*availability*), how cost-effective those resources are in generating value for the focal organization (*productivity*), and how much the cost-effectiveness of those resources is limited to a narrower set of tasks given their organizational context (*specialization*). The instructor can capture these relationships as illustrated in *Board #1*. X's workforce, as is often the case in R&D, was one that is relatively small, highly productive, and relatively specialized. Thus, X was at risk of paying large opportunity costs.

TN Figure 4 Distribution of process outputs



Source: Case author.

Beyond these dimensions, it is useful to show how the high-variance strategy of X, which generated a highly skewed distribution of outcomes, is connected to opportunity costs. To do this, the instructor can draw **TN Figure 4** as *Board #2*. First, the instructor should emphasize how standard operations management focuses on minimizing variance and generating distributions of output that look like the "standard" line. However, when the manager only needs to choose a very small, perhaps a single, output then the manager will be willing to produce more low-value output in order to have a chance at producing a very high-value output per the "innovative" line. To put additional emphasis on this point, the instructor can draw a third line to represent X that illustrates an extreme view of almost entirely low-quality value, but a very small chance of exceptionally high-value output.[h]

[h] For comedic effect it is fun to continue X's line on the distribution as far as the board will allow the instructor to, possibly onto another board in the classroom.

After this, the instructor should provide the class with a brief wrap of the first lesson of the case:

> **Lesson**: *Opportunity costs loom large for innovative processes, so it is essential to be thoughtful about which components of a project are pursued first.*

Then, the instructor can use the following question to build a bridge to the next part of class:

> **Q:** *I'd like to hear some stories about projects you worked on that weren't killed soon enough. Who would like to share an example?*

It is helpful to push students to hypothesize the specific reasons why the project was not killed, the source of that friction, and their sense of the opportunity costs paid because of this mistake. This often yields stories centered on misaligned incentives, the "sunk-cost fallacy", a notion of making progress just for progress's sake, or asymmetric information.

3. *X's operating model: Incentives & culture (10 minutes)*

To introduce the next discussion, the instructor can build on the students' stories with the following question:

> **Q:** *It sounds like many of you could have used more kill metrics in your prior positions. But maybe that wouldn't be enough. At X, kill metrics and milestones were just two features of their operating model that facilitate their mission. What other features of X's operating model stood out to you?*

The instructor should collect students' answers to this question on *Board #3* (**TN Appendix A**). For each answer, it is useful to ask students about the downsides of each feature, especially if taken to the extreme. While collecting these features, it is engaging to ask students how they would feel if they worked at an organization that had these sorts of features. This highlights the challenge of sustaining an organization with the sort of culture that X had, and it motivates the following question:

> **Q:** *If you were the lead recruiter for X, how would you recruit new employees? How would you interview applicants?*

This question usually yields a wild range of ideas, all of which emphasize the importance of selecting individuals who are tolerant of risks, ambiguities, failures, and changing the nature of their work very rapidly. The instructor should wrap this part with a lesson that connects back to the distribution of ideas from *Board* #2 with issues of culture and incentives:

> **Lesson:** *Look back at what X was trying to do: generate a lot of "crazy" ideas that would most often fail but were in hopes of finding a very small number of enormously valuable ideas; they cared about maximums, not averages. The challenge is that most people are not built to think that way—we typically evaluate ourselves and each other based on concepts like "how good are we on average?" That was incompatible with X's strategy. So, all of these design choices X was making (i.e., a separate building from Google, failure parties, "passionate dispassion") were in service of making people comfortable with the pursuit of maximums, not averages, but it's hard!*

If time permits, one particularly fun question (especially for use in a course where classroom participation comprises a large fraction of students' grades) that can emphasize the uniqueness of evaluating people based on their maximum performance is the following:

> **Q:** *How do you think the comments in our classroom would change if I explicitly said at the beginning of the semester that I would grade you based on the quality of your single best comment?*

Students' answers to this question vary widely but they quickly see that this would dramatically affect their classroom participation strategies, likely leading them to try many more different ways of contributing to the discussion.

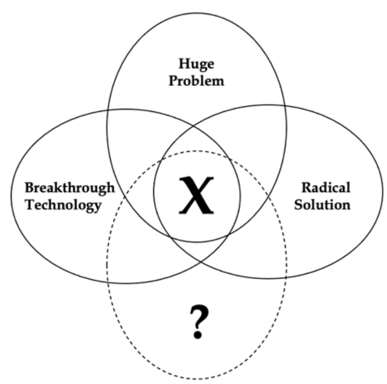
4. *Mission statement and missing ideas (15 minutes)*

To bridge into this next part of the class, the instructor can note that X's mission statement loomed above all of the specific features of their operating model we just discussed, and then the instructor can draw the three circles for X's criteria of Huge Problem, Radical Solution, and Breakthrough Technology on *Board #4* (**TN Figure 5**, sans the fourth circle with the question mark). Then the instructor can ask the following:

> **Q:** *Building off of our understanding of X's operating model, and with their mission in mind—what sorts of ideas do you think X might be missing out on?*

Students' answers to this question are usually specific examples of ideas that meet only one or two of the three criteria. For example, X may miss problems that are meaningful in size but not so grand as to be on the scale of another Google-level company, or X may miss solutions that are extremely cost-effective and make use of very simple technologies that have simply not been combined in a certain way previously.

TN Figure 5 X's mission statement possibility



Source: Case author.

Frequently, many students will argue for the removal of one or more of the criteria. This provides a nice motivation for the following question which the instructor can ask while drawing a fourth circle with a question mark on the board:

> **Q:** *Many of you seemed to be in favor of relaxing X's mission statement to be less constrained. But now, let's imagine you're Sundar Pichai, the CEO of Alphabet, and Astro Teller has asked you to be more specific about X's mission. Specifically, you get to add one more criterion. This additional criterion won't be made public, it's just to help guide Teller's decisions, so let's not worry about any public relation concerns of how this additional criterion is phrased. What would you want the fourth circle to be?*

Examples of students' answers, which the instructor should collect around the question mark, are shown on *Board #4* (**TN Appendix A**). For each response, it is helpful to push students to be clear about why they chose that particular new criterion, what sort of ideas might have been missed without that criterion, and whether the student thinks that criterion might already be implicitly understood by Teller. After a bit, it is useful to push towards a learning point with the following question:

> **Q:** *Who cares about this? These are just words; do we think they are actually that important?*

Framing this question in a provocative way will generate at least one response from a student who agrees that it does not matter. After such a comment, the instructor can refer to the quote mentioned in the *Case Analysis* regarding the high-profile Google employee moving to OpenAI because of the mission, which is copied here (the following is an excerpt from a news article covering competition between OpenAI and Google for talent; includes a quote of Ilya Sutskever, who left Google and became the Chief Scientist at OpenAI):

> *"They did make it very compelling for me to stay, so it wasn't an easy decision," Sutskever says of Google, his former employer. "But in the end, I decided to go with OpenAI, partly of because of the very strong group of people and, to a very large extent, because of its mission."*[17]

This allows the instructor to not debate *whether* mission statements are valuable, but rather *how and when* they are valuable. This discussion can continue for some time, after which the instructor can then wrap with the following lesson:

> **Lesson:** *Mission statements provide "soft" incentives for employees engaged in processes where inputs and outputs are hard to observe, and employees derive some value from having agency over their pursuits and the specifics of what they work on.*

5. *Evaluating innovative processes (15 minutes)*

To motivate the next part of the class, the instructor can ask the following:

> **Q:** *Now that we understand how X's mission could shape which projects were pursued, and how kill metrics and other incentives shaped how those projects were pursued. I want to discuss how evaluation should happen at X. How do you think we should evaluate the people of X? Or X as a whole? Should X have its own, organization-level kill metric?*

Students can struggle a bit with this question and may offer answers that are based on their perceptions of how valuable X's observable work to date has been. It is useful to pose the following question to push the students further:

> **Q:** *What is different about evaluating an X employee compared to, say, an average employee at Google or other similar organizations?*

This will help students realize that, at X, there was no customer providing feedback, it was very difficult to monitor employees' efforts, it was very difficult to judge their output in the near-term, etc. If any students have work experience with Google or know of Google's OKR (Objective and Key Results) processes, you can ask them to elaborate on how that evaluation process works and to consider how something like that might have worked (or not) at X—noting that X does not use OKRs.

Overall, the goal of this discussion is to help students realize that the input and outputs of processes at X were very difficult to observe, are rare, and involved a large degree of uncertainty, which has important implications for how evaluations should be conducted. In the extreme, it suggests that all X could do is rely on a theory of innovation

Making an analogy to gambling can be quite instructive (i.e., positive expected value bets do not always pay off ex-post). To summarize this point while contrasting it with standard process analyses, the instructor can construct **TN Figure 2** on *Board #5* (**TN Appendix A**) and wrap this part with the following lesson:

> **<u>Lesson</u>:** *Processes designed for innovation often involve inputs that are hard to observe and outputs that are rare and full of chance. In this setting, simulations and experiments may not be feasible evaluation tools. Here, managers need to rely on their theories of how the innovative processes works. As Teller has often put it in his public remarks: he evaluated X based on "habits, not outcomes." He knew that judging any particular person or team, or project, or even all of X, purely based on its outcomes was a lot like judging whether a bet was a good idea based on whether you win the bet (which you shouldn't!). Rather, he understands that you should judge a bet based on whether it was a good idea in expectation, or "in theory." How do you develop your own theories? By engaging with research on how innovation works, by reading historical accounts of innovators and innovative organizations, and by having discussions like this these we're having in this class right now!*

6. *Foghorn debate & learning curves (20 minutes)*

The instructor can transition into the debate about Foghorn with the following question:

> **Q:** *Now that we have a good, mutual understanding of X, let's return to the key matter at hand: Foghorn. I'm guessing many of you have already seen the decision X actually made; they killed it. But Teller has said publicly that this project was one of the toughest kills they have made. So, let's re-debate that decision. Should X kill Foghorn or not?*

X was very public about their decision to kill Foghorn, which is why the instructor can break the norm of keeping the discussion void of post-case facts and pose the question this way (note: it is very easy to fine online coverage of X's decision to kill foghorn). For students who argue Foghorn should be continued, the instructor can ask the following to test their assumptions and beliefs about potential competitors:

> **Q:** *How do you think Alphabet will be able to capture the value generated by Foghorn? Do you expect them to compete with large oil producers (e.g., OPEC)? How does your forecast about battery-powered vehicles play in here?*

For students who argue Foghorn should be killed, the instructor can ask the following to test their understandings of X's mission and the options it has when killing projects:

> **Q:** *Are you worried about losing one of X's largest projects in the clean energy space? There is still roughly 2 years before the deadline of Hannun's kill metric, why kill it now? Isn't this X's opportunity to show it's not just "AI and robots"? How exactly would you kill the project and what would you do with the technology and developed so far?*

At some point during the debate, a student may reference Exhibit 10 of the case and the continual rate of progress in the efficiency of the technology. If not, the instructor can inject this into the conversation to see students' thoughts. As this discussion about progress begins, the instructor can bring up the slide that shows the Foghorn learning curve more clearly (**TN Figure 3**, excluding the OpenAI learning curve initially).

After permitting students some time to digest and debate the Foghorn learning curve, the instructor can then tell the students that they'd like to show them another learning curve for a comparison. Here, the instructor should reveal and explain the OpenAI learning curve and pose a question:

> *Managers of R&D projects often have views of this sort* [referencing the Foghorn curve], *which are referred to as "learning curves" that plot progress on the y-axis and some other variable, such as time or investment, on the x-axis. To give you another example, I want to share a learning curve from the company OpenAI, who released a paper in 2020 that showed this particular learning curve for one of their neural language models* [here, the instructor should reveal the OpenAI curve]. *The measure of progress on the y-axis of the OpenAI is similar to the efficiency measure on the Foghorn curve; smaller is better. But the x-axis is different. It is the size of the dataset used to train the models.*
>
> *For context, this learning curve was probably what the managers, scientists, and engineers at OpenAI were looking at when they were debating the transition from GPT-2 to GPT-3, which was expected to cost them more than two orders of magnitudes of additional costs, going from tens of thousands to millions of dollars.*
>
> **Q:** *With our benefit of hindsight, why do we think X looked at their learning curve and said "Kill" while OpenAI looked at their learning curve and said "Onward"?*

There are many reasonable answers to this question. But the point is to get students to realize that the Foghorn learning curve does not represent a managerially relevant causal relationship, while the OpenAI curve does. Yes, Foghorn's curve suggests to the manager that continued work on the project may continue to yield progress, but it suggests no specific aspect or component of the project where further investment would *cause* progress. Conversely, the OpenAI learning curve represents a relationship that clearly suggests that increasing the training dataset size will yield improvement in the model. The instructor can summarize this point with the following:

> **Lesson:** *Patterns in progress, as illustrated in learning curves, are more valuable the more that they reflect a causal relationship between a specific aspect of a project that the manager can influence and the measure of progress.*

In order to move the class towards a final summary, the instructor can conduct another poll of the students to see, after our discussion, if they think Foghorn should be killed or not. If time permits, it useful to perform two polls here. In the first, the instructor can force the students to take the perspective of Teller, to truly consider his incentives and the position of X within Alphabet, and to ignore any value created by Foghorn that might not be captured by Alphabet. This helps students set aside notions of social value that may have arisen earlier in the class, a focus primarily on X and Alphabet as stakeholders. Then, in the second poll, the instructor can ask students whether, from society's perspective, Foghorn should be killed or not. Unsurprisingly, these two polls will often yield a majority for "Kill" in the first and a majority for "Continue" in the second. This motivates much larger questions about private and social value of R&D that are beyond the scope of this case but are elevated questions that provide a thought-provoking end to the class.

7. *Summary (5 minutes)*

To begin the summary, it is useful to emphasize the challenge of working in this sort of environment. One way of starting this discussion is to ask:

> **Q:** *Let's take a step back from Foghorn and reconsider X as a whole. It seems like a pretty wild place to work. It makes me wonder, assuming you had the technical skills required, would you want to work at X?*

After receiving some students' thoughts on this question, which would ideally include both some enthusiastic and pessimistic responses, it may be useful to share the following anecdote from Obi Felton, who worked at the X at the time of the case as the "*Head of Getting Moonshots Ready for Contact*

With the Real World," but then departed the organization afterwards. It is an excerpt from the *Freakonomics* podcast with Obi, and is an exchange between her and the host (Steven Levitt):[18]

> *LEVITT: When I heard about your job at X, it sounded like a dream job to me. And then the way you've spoken about it so far, it seems like it was a dream job for you also. But you left a few years ago to do something else. How come?*
>
> *FELTEN: I'd say my job at X was a dream job, for sure for the first few years. But it was actually really soul destroying for me. It was this endless cycle of just launching projects, hiring people, and then having to fire them again when we shut the project down. And it took a toll not just on the teams, but also on me, which I now realize looking back. The pressure of producing the next self-driving car and the next balloon project was absolutely immense, like it couldn't just be any old thing. It had to be something really incredibly amazing that was also very undefined. Whereas when I worked at marketing at Google, I had really clear metrics. Like we get 100 million Chrome users. That's successful. If we get only a million, that's not as successful. But at X everything was just a little fuzzier and less measurable. And then the final thing was a good friend and coach told me, "You are just in the wrong job for you because there's some people who are scuba divers, they love to go deep, and there's some people who are snorkelers, who love to go across a whole bunch of things. And you're a scuba diver in a snorkeling job." … I had all this diversity in the portfolio, which meant I could never go deep on anything. And it wasn't my job. It was the team's job to go deep on them. So, I suddenly realized, oh, this is a dream job for other people, but it's not a dream job for me. I've been happiest when I've been building things. So I went to Astro, the head of X, and I said, "Look, I'm going to quit. I'm going to go back to the startup world." And he said, "What? Like, why would you want to leave?" And I said, "Because I want to build something again. You know, I'm a builder at heart."*

After sharing this quote, or some abbreviated version, it is useful to ask for further student reflections on the culture and incentives of X. If time permits, it is helpful to allow students to reflect more broadly on the discussion and provide the class with their own key takeaways from the discussion.

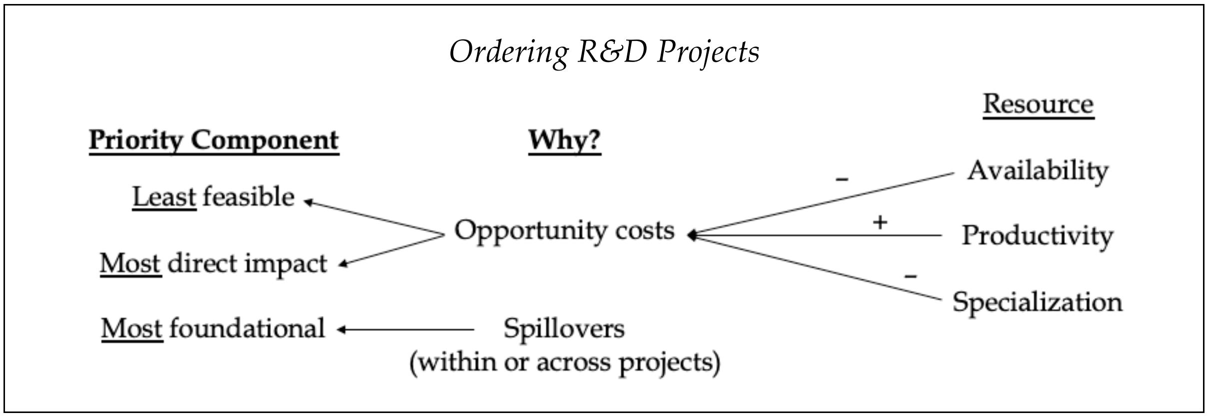
> **Q:** *More generally, what open questions do you think remain? What have you learned from X and the Foghorn project that will be useful going forward in your work more generally?*

The instructor can then wrap the discussion with a summary along the following lines:

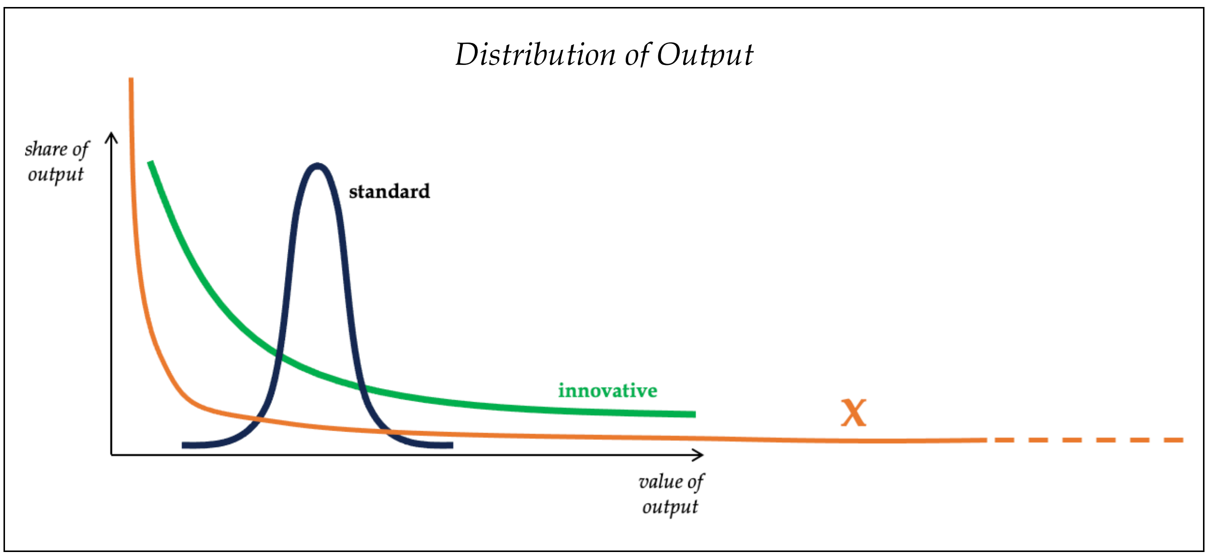
> *Killing fast isn't just a slogan—X clearly made specific operational choices and focused on a very specific culture to ensure projects were killed fast. Why? Because opportunity costs loom large in this world of extreme values* [reference Boards #1-2]. *But killing fast is not an easy strategy to implement, nor is it universally the right approach—in settings where the foundational value of work on certain ideas is large, succeeding slowly might be more efficient. In both cases, accurately forecasting progress requires an understanding of what is causing progress, not just what is correlated with it* [reference the learning curve slide].
>
> *And it certainly may be the case that X was missing out on some "great" ideas because of the way they set their mission statement and the specific culture and incentives they had put into place* [reference Boards #3-4]. *But Teller clearly had a theory of how innovation works and, in the world of innovation where inputs and outputs are often very hard to see,* [reference Board #5] *we should be careful to judge a theory based on its output. Rather we should judge the theory itself, but that is very hard to do!*

TN Appendix A —Board Plan

Board #1



Board #2

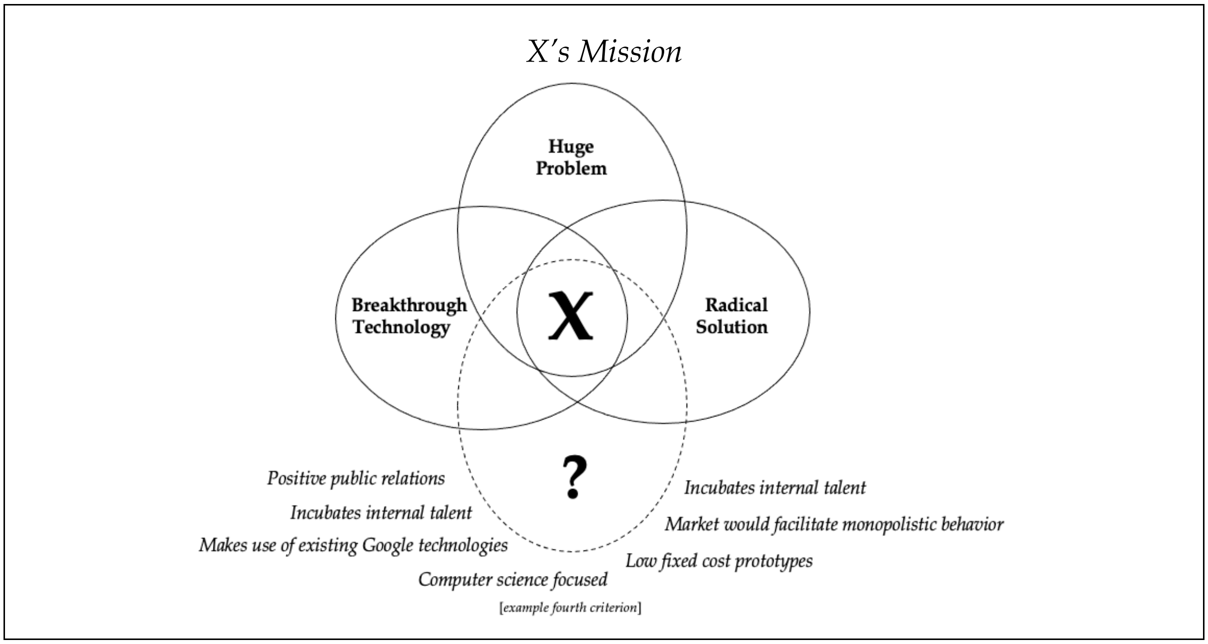


Board #3

X's Operating Model

- Culture, "dispassionate – passion", psychological safety
- Recruitment: risk-tolerant, jack-of-all-trades, academics (temporarily)
- No equity in early success
- Directly incentivize failure (e.g., bonuses, failure parties)
- Physical separation of X from Google HQ
- "Foundry" to theorize about business model

Board #4



Board #5

Evaluating Innovative Processes

Evaluation method:	**theories**	**experiments**	**simulations**
Key metric:	**inputs**	**proxies**	**outputs**

↔

low (innovative) — *process observability & certainty* — *high (standard)*

Endnotes

[1] Source: https://blog.x.company/a-peek-inside-the-moonshot-factory-operating-manual-f5c33c9ab4d7#.je9v63r20.

[2] Source: https://blog.x.company/tips-for-unleashing-radical-creativity-f4ba55602e17.

[3] See: Reagans, R., & Zuckerman, E. W. (2001). Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organization Science*, *12*(4), 502-517.

[4] Source: https://www.wired.com/2013/02/moonshots-matter-heres-how-to-make-them-happen/.

[5] See: Manso, G. (2011). Motivating innovation. *Journal of Finance*, *66*(5), 1823-1860; Halac, M., Kartik, N., & Liu, Q. (2016). Optimal contracts for experimentation. *Review of Economic Studies*, *83*(3), 1040-1091; Guo, Y. (2016). Dynamic delegation of experimentation. *American Economic Review*, *106*(8), 1969-2008; Halac, M., Kartik, N., & Liu, Q. (2017). Contests for experimentation. *Journal of Political Economy*, *125*(5), 1523-1569; Khalil, F., Lawarree, J., & Rodivilov, A. (2020). Learning from failures: Optimal contracts for experimentation and production. *Journal of Economic Theory*, *190*, 105107; Dmitriev, D. (2023). Motivating creativity, *Mimeo*.

[6] See: Brown, J. S. (1991). Research that reinvents the corporation. *Harvard Business Review*, *69*(1), 102-110; Kelley, R., & Caplan, J. (1993). How Bell Labs creates star performers. *Harvard Business Review*, *71*(4), 128-139; Gertner, J. (2012). *The Idea Factory: Bell Labs and the Great Age of American Innovation*. Penguin: London.

[7] See: Granot, D., & Zuckerman, D. (1991). Optimal sequencing and resource allocation in research and development projects. *Management Science*, *37*(2), 140-156; De Reyck, B., & Leus, R. (2008). R&D project scheduling when activities may fail. *IIE Transactions*, *40*(4), 367-384.

[8] See: Myers, K. R., & Lanahan, L. (2022). Estimating spillovers from publicly funded R&D: Evidence from the US Department of Energy. *American Economic Review*, *112*(7), 2393-2423.; Bloom, N., Schankerman, M., & Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, *81*(4), 1347-1393.

[9] See: Myers, K. R. (2020). The elasticity of science. *American Economic Journal: Applied Economics*, *12*(4), 103-134.

[10] See: Aghion, P., Dewatripont, M., & Stein, J. C. (2008). Academic freedom, private-sector focus, and the process of innovation. *RAND Journal of Economics*, *39*(3), 617-635.

[11] See: Chapter 2 of Edmondson, A. C. (2023). *Right Kind of Wrong: The Science of Failing Well*. Simon and Schuster: New York City.

[12] See: Edmondson, A. C. (2011). Strategies for learning from failure. *Harvard Business Review, 89*(4), 48-55; Edmondson, A. C. (2023). *Right Kind of Wrong: The Science of Failing Well*. Simon and Schuster: New York City; Hill, L. A., Brandeau, G., Truelove, E., Lineback, K. (2014). *Collective Genius: The Art and Practice of Leading Innovation*. Harvard Business Review Press: Boston.

[13] See: Akerlof, G. A., & Kranton, R. E. (2005). Identity and the economics of organizations. *Journal of Economic Perspectives*, *19*(1), 9-32; Bolton, P., Brunnermeier, M. K., & Veldkamp, L. (2013). Leadership, coordination, and corporate culture. *Review of Economic Studies*, *80*(2), 512-537.

[14] Source: https://www.wired.com/2016/04/openai-elon-musk-sam-altman-plan-to-set-artificial-intelligence-free/.

[15] Source: https://www.macrotrends.net/stocks/charts/GOOGL/alphabet/cash-on-hand.

[16] See: Weitzman, M. L. (1998). Recombinant growth. *Quarterly Journal of Economics*, *113*(2), 331-360.

[17] Source: https://www.wired.com/2016/04/openai-elon-musk-sam-altman-plan-to-set-artificial-intelligence-free/.

[18] Source: https://freakonomics.com/podcast/can-a-moonshot-approach-to-mental-health-work.