A Quest for Knowledge

Christoph Carnehl Johannes Schneider

February 2023

Motivation

In his 1945 letter to Roosevelt—*Science, the Endless Frontier*—which paved the way for the creation of the NSF, Vannevar Bush emphasizes the value of research for society and the importance of scientific freedom.

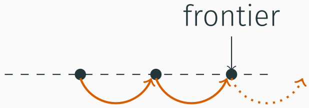
But...

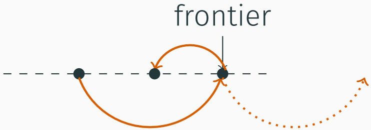
- How do researchers act under scientific freedom?
- What are the implications for the evolution of knowledge?
- Is there any merit to focus funding on “super novel” research?

Contribution

1. Set up flexible framework that connects
    - novelty of research
    - research output
    - the value to society
2. Identify a simple static friction

    → cost of effort harms both novelty and output.
    - consistent with empirical literature (Rzhetsky et al., 2015)
3. Dynamic implications:
    - no-cost benchmark: ladder structure
    - laissez-fair: ladder structure
    - constrained optimal: research cycles





Agenda

1. (static) Model
2. (static) Benefits and Cost
3. (static) Researcher's Choice
4. The Evolution of Knowledge
5. Literature

Model

Model Basics

2 players, Researcher and Decision Maker

1. R observes initial knowledge, $\mathcal{F}_k$.
2. R selects a question, $x \in \mathbb{R}$, and a research intensity, $\rho \in [0, 1]$.
3. If R obtains a discovery $y(x)$, knowledge is augmented by it, $\mathcal{F}_k \cup (x, y(x))$.
4. DM observes current knowledge and selects $a(\cdot)$ to address many problems.

Truth, Knowledge, and Research Areas

Questions: Each $x \in \mathbb{R}$ is a question.

Answers: The answer to x is the realization $y(x) \in \mathbb{R}$ of a random variable $Y(x)$.

Truth: The realization of a standard Brownian path determining all $y(x)$.

Knowledge: Set of known question-answer pairs

$$\mathcal{F}_k = \{(x_1, y(x_1)), \ldots, (x_k, y(x_k))\}, \text{ with } x_1 < x_2 < \ldots < x_k.$$

$\Rightarrow$ Knowledge partitions questions into **research areas**

$$\{\underbrace{(-\infty, x_1)}_{\text{area } 0}, \underbrace{[x_1, x_2)}_{\text{area } 1}, \cdots, \underbrace{[x_{k-1}, x_k)}_{\text{area } k-1}, \underbrace{[x_k, \infty)}_{\text{area } k}\}.$$

Research area i has **length** $X_i := x_{i+1} - x_i$.

Conjectures

A **conjecture** is the distribution of the answer $y(x)$ to a question x: $G_x(Y|\mathcal{F}_k)$.

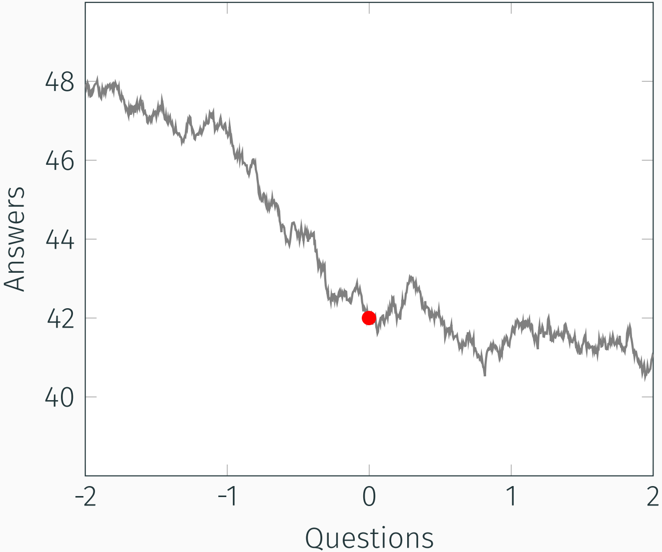
$\Rightarrow$ Brownian path determines answers: $Y(x|\mathcal{F}_k) \sim \mathcal{N}\left(\mu_x(Y|\mathcal{F}_k), \sigma_x^2(Y|\mathcal{F}_k)\right)$ with

$$\mu_x(Y|\mathcal{F}_k) = \begin{cases} y(x_1) & \text{if } x < x_1 \\ y(x_i) + (x - x_i)\frac{y(x_{i+1}) - y(x_i)}{X_i} & \text{if } x \in [x_i, x_{i+1}) \\ y(x_k) & \text{if } x \geqslant x_k \end{cases}$$

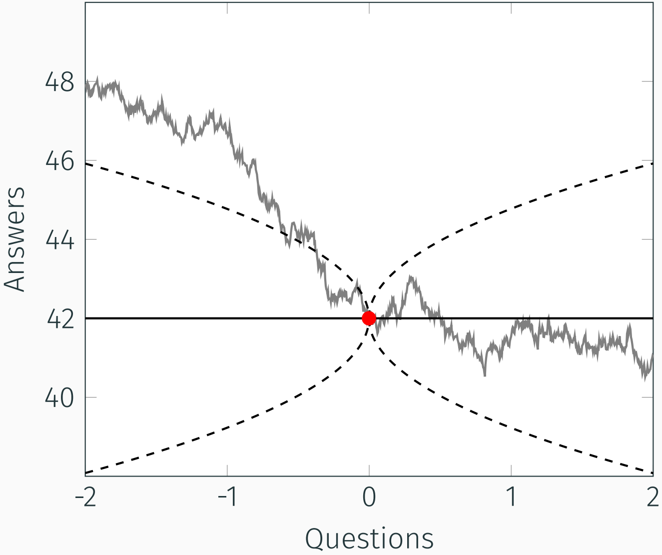
$$\sigma_x^2(Y|\mathcal{F}_k) = \begin{cases} x_1 - x & \text{if } x < x_1 \\ \frac{(x_{i+1} - x)(x - x_i)}{X_i} & \text{if } x \in [x_i, x_{i+1}) \\ x - x_k & \text{if } x \geqslant x_k. \end{cases}$$

Model of Knowledge - Graphically

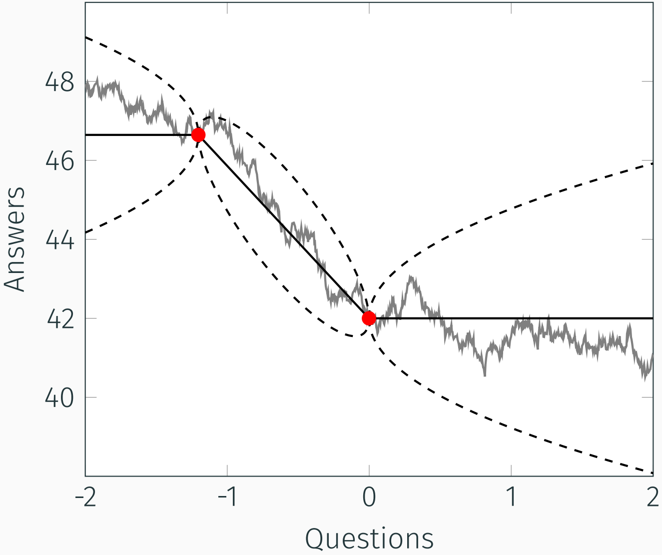
Truth and Knowledge

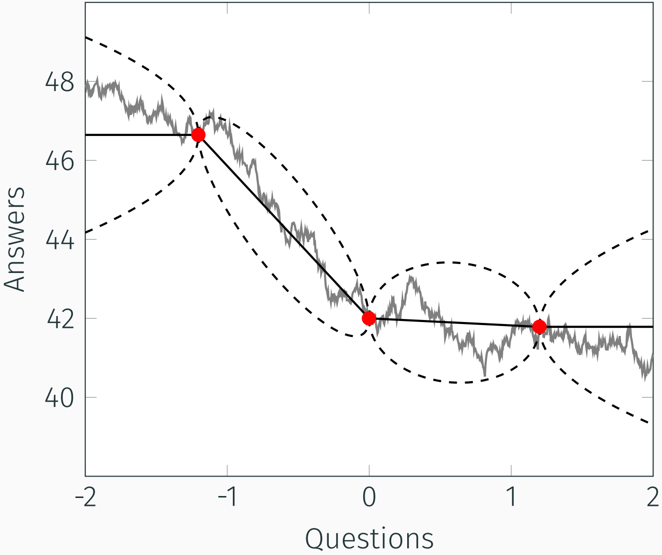


Conjectures

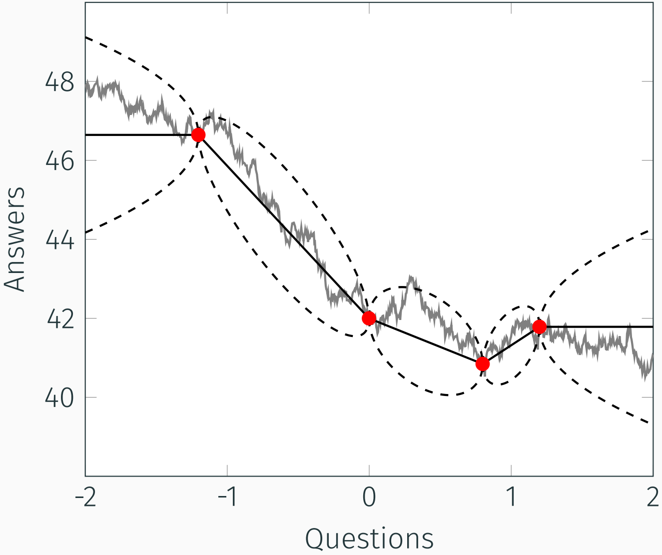


Expanding Knowledge




48
46
44
42
40
Answers
-2
-1
0
1
2
Questions

Deepening Knowledge



Society as Decision Maker

Decision Making

Society, represented by DM, observes $\mathcal{F}$ and makes decisions on *all* questions.

For each question x, she can

- stick with the status quo: $a(x) = \varnothing$ or
- make a proactive choice: $a(x) \in \mathbb{R}$

with per-question payoffs

$$u(a(x), x) = \begin{cases} 0 & \text{if } a(x) = \varnothing, \\ 1 - \frac{(a(x)-y(x))^2}{q} & \text{if } a(x) \in \mathbb{R}. \end{cases}$$

Status quo guarantees a finite payoff but proactive choices can be beneficial if sufficiently good—error tolerance of $\sqrt{q}$.

Benefit of Discovery

What is the Value of Knowledge?

Jacob Marschak (1974):

Knowledge is useful if it helps to make the best decisions.

Hjort, Moreira, Rao and Santini (2021):

- science fosters the adoption of effective policies and
- more precise information improves policies further.

The Value of Knowledge

Knowledge benefits society via more precise conjectures about optimal policies.

$$a^*(x|\mathcal{F}) = \begin{cases} \mu_x(Y|\mathcal{F}) & \text{, if } \sigma_x^2(Y|\mathcal{F}) \leqslant q \\ \varnothing & \text{, if } \sigma_x^2(Y|\mathcal{F}_k) > q. \end{cases}$$

Only if society's conjecture about the answer is sufficiently precise, a proactive choice is optimal.

Society's value of knowledge is

$$v(\mathcal{F}) := \int_{-\infty}^{\infty} \underbrace{\max\left\{1 - \frac{\sigma_x^2(Y|\mathcal{F})}{q}, 0\right\}}_{=u(a*(x),x)} dx.$$

Benefit of a Discovery

The discovery of an answer $y(x)$ to question x enhances knowledge to

$$\mathcal{F}_k \cup (x, y(x)).$$

The benefit of a discovery is how it improves decision making

$$V(x; \mathcal{F}_k) := v\left(\mathcal{F}_k \cup \Big(x, y(x)\Big)\right) - v\left(\mathcal{F}_k\right).$$

x_1 and x_k are the frontiers of knowledge. A discovery

- expands knowledge if $x \notin [x_1, x_k]$ and
- deepens knowledge if $x \in [x_1, x_k]$.

Change of Variables

We can simplify by focusing on

- the distance to knowledge, $d(x; \mathcal{F}_k) := \min_{\xi \in \{x_1, \ldots, x_k\}} |x - \xi|$
- the length of the research area in which x lies, X.

Applying this rewriting to the variance,

$$\sigma^2(d; X) := \sigma_X^2(Y|\mathcal{F}_k) = \frac{d(X-d)}{X}.$$

Note that for expanding knowledge

$$\sigma^2(d; X = \infty) = d.$$

$\Rightarrow$ Benefit of discovery $V(d; X)$ determined by *d*istance and length of area, X.

Benefit of Discovery - Characterization

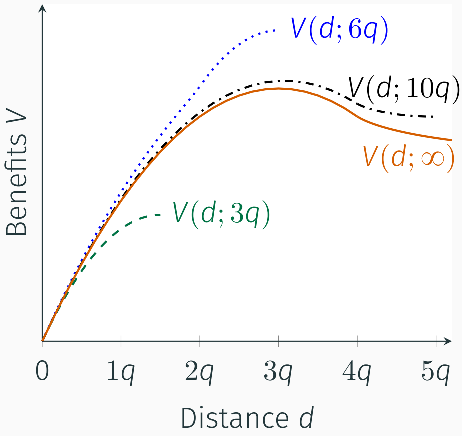
Can fully characterize (in the paper)

Main points:

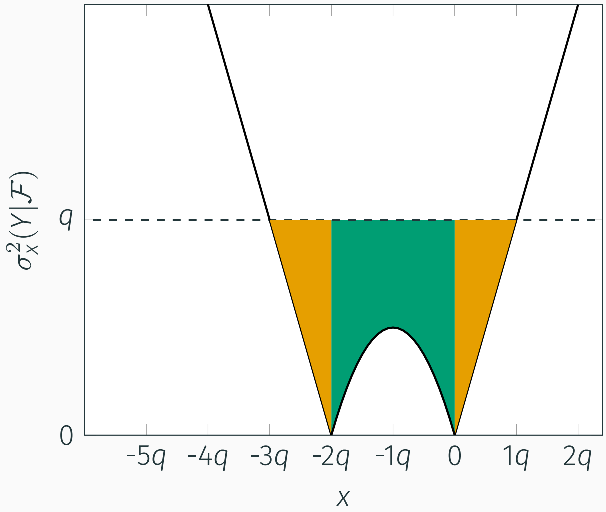
- $V(d;X)$ linear in q
- Given X, increasing in d first

  for (X,d) large decreasing; *max* $\geqslant 3q$
- Given d *optimal*,

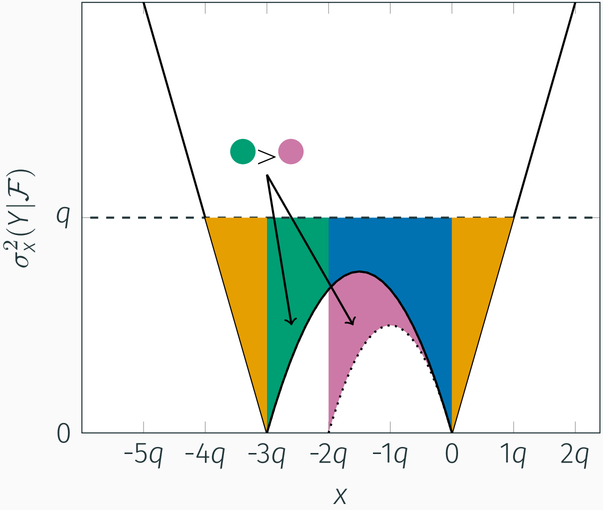
  non-monotone in X with interior max
- For $X=\infty$, optimal $d=3q$.



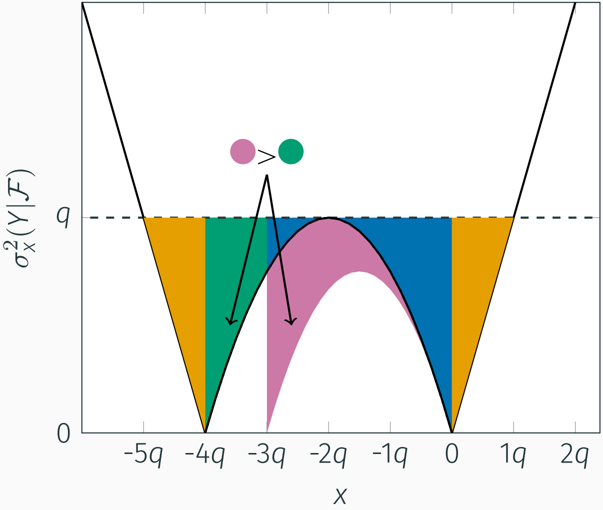
Benefit of Expanding Knowledge



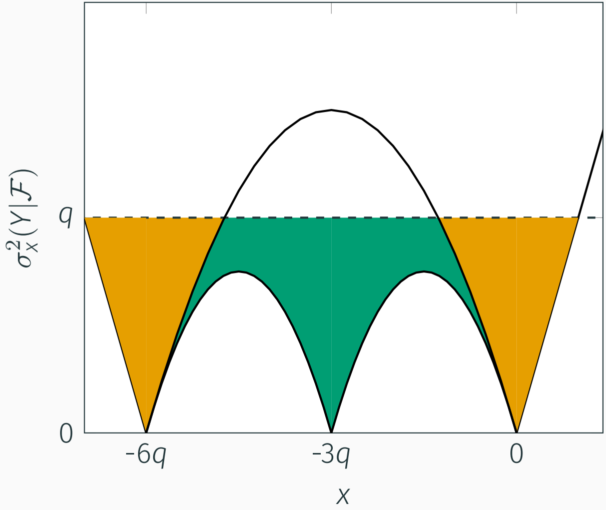
Benefit of Expanding Knowledge



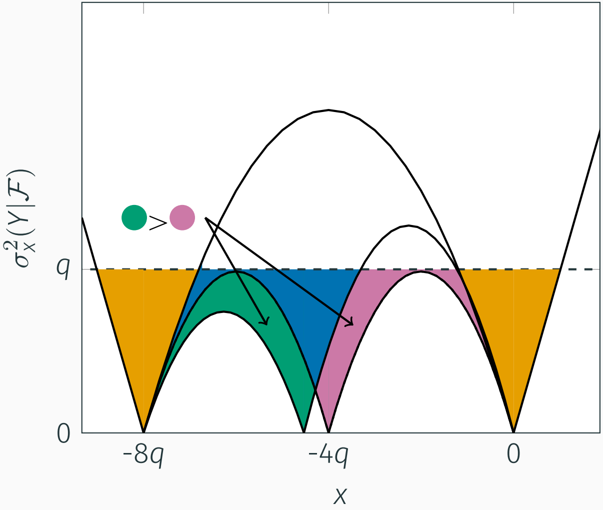
Benefit of Expanding Knowledge



Deepening Knowledge



Deepening Knowledge



Benefit-Maximizing Distance

Corollary

The benefit-maximizing distance $d^0(X)$ in a research area of length X has the following properties:

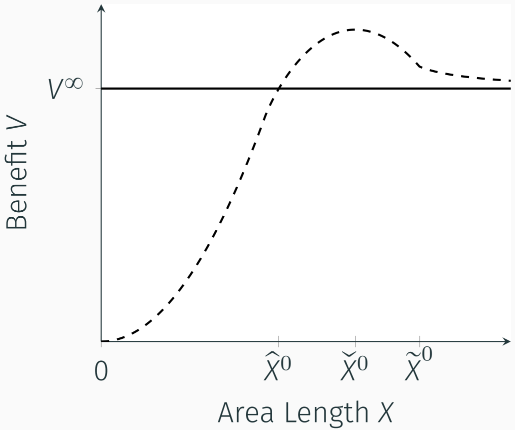
- *If $X = \infty$, $d^0(\infty) = 3q$.*
- *If $X \leqslant \widetilde{X}^0 \in (6q, 8q)$, $d^0(X) = X/2$.*
- *If $X \in (\widetilde{X}^0, \infty)$, $d^0(X) \in (3q, X/2)$.*
- *$d^0(X)$ is increasing in X for $X < \widetilde{X}^0$ and decreasing for $X > \widetilde{X}^0$.*

Properties of Benefit of Discovery

Corollary

2 cutoffs $4q < \widehat{X}^0 < 6q < \check{X}^0 < 8q$, *s.t.*

- *benefit of expanding knowledge by* $3q$ ***dominates*** *iff* ***all*** $X_i < \widehat{X}^0$.
- *benefit of deepening knowledge*
  - $\uparrow$ *in* X *if* $X < \check{X}^0$
  - $\downarrow$ *in* X *if* $X > \check{X}^0$.



Cost of Research

Research as Search for an Answer

The researcher searches for an answer $y(x)$ by sampling an interval $[a, b] \subseteq \mathbb{R}$.

The researcher discovers the answer $y(x)$ iff $y(x) \in [a, b]$.

Searching for an answer is costly: $c([a, b]) = \eta(b - a)^2$.

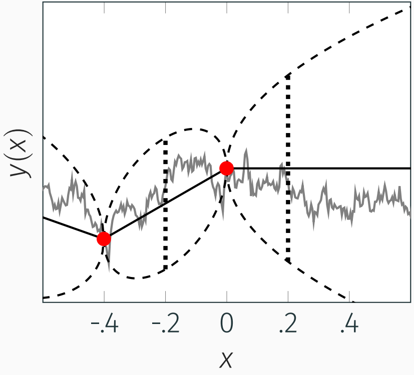
Lemma

Given a question x with distance d in a research area of length X, the lowest-cost search interval such that the answer is contained in the interval with probability ρ has cost

$$\eta c(\rho, d; X) = 8\eta \big(erf^{-1}(\rho)\big)^2 \sigma^2(d; X).$$

Cost of Research Grpahically

Suppose you want to obtain an answer with probability 95%



The right interval is $1 + (\sqrt{2} - 1)/\sqrt{2}$ (≈ 1.3) times the left interval.

Researcher’s Choice

How to Choose Research Questions?

Biologist and Nobel laureate Peter Medawar (1976):

Research is surely the art of the soluble. (...) Good scientists study the most important problems they think they can solve.

Researcher's Decision Problem

Researcher stands on shoulders of giants and observes $\mathcal{F}_k$.

Researcher's payoff consists of the benefit of discovery and the cost of search.

Researcher decides on a research question $x \in \mathbb{R}$ and a search interval $[a, b] \subseteq \mathbb{R}$.

The choice of x and $[a, b]$, can be reduced to a choice of

- a research area denoted by its length, X,
- a distance to existing knowledge, d,
- a success probability of search, ρ.

$$\max_{X \in \{X_0, \ldots, X_k\}} \underbrace{\max_{\substack{d \in [0, X/2], \\ \rho \in [0,1]}} \rho V(d; X) - \eta c(\rho, d; X)}_{=: U_R(X)}$$

Optimal Choice: Distance, Novelty and Research Area

Proposition

Suppose $\eta > 0$. There is $\widehat{X} \leqslant \dot{X} \leqslant \check{X} \leqslant \widetilde{X} < 8q$ s.t.:

- *R expands iff all X shorter than $\widehat{X}$.*
- *R's payoffs, $U_R(X)$, are single peaked with a max at $\check{X}$.*
- *Optimal distance, $d^\eta(X)$, and prob. of discovery, $\rho^\eta(X)$, are non-monotone in X.*
- *$\rho^\eta(X)$ has its max at $\dot{X}$*
- *$d^\eta(X)$ at $\widetilde{X}$.*

pictures

Researcher: Main Take-Away

Novelty choice d^η, choice of X and payoffs qualitatively similar to society's choices.

Output choice ρ^η largest for intermediate areas.

Substitutes or Complements?

The Evolution of Knowledge

Dynamic Game

Discrete time $t = 1, 2, ...$

Each period a (short-lived) DM plays with a short-lived R taking $\mathcal{F}_t$ as given

Society's payoff is discounted sum of DM payoffs.

$$\mathbb{E}\left[\sum_{t=1}^{T} \delta^{t-1} v(\mathcal{F}_{t+1})\right].$$

Assumption

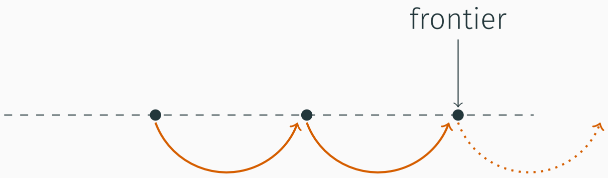
- *All R same cost type η.*
- *R conditions only on current knowledge, $\mathcal{F}_t$.*
- *Symmetric pure strategies.*

A few sanity checks

Corollary

Without interference and independent of $\mathcal{F}_1$ researchers will

- *first close gaps until* $\max X < \hat{X}$,
- *then aim at pushing the frontier step-by-step with* $(\rho^\eta(\infty), d^\eta(\infty))$,
- *eventually fail to improve knowledge*



Thought Experiment

Suppose there is an initial period $t = 0$

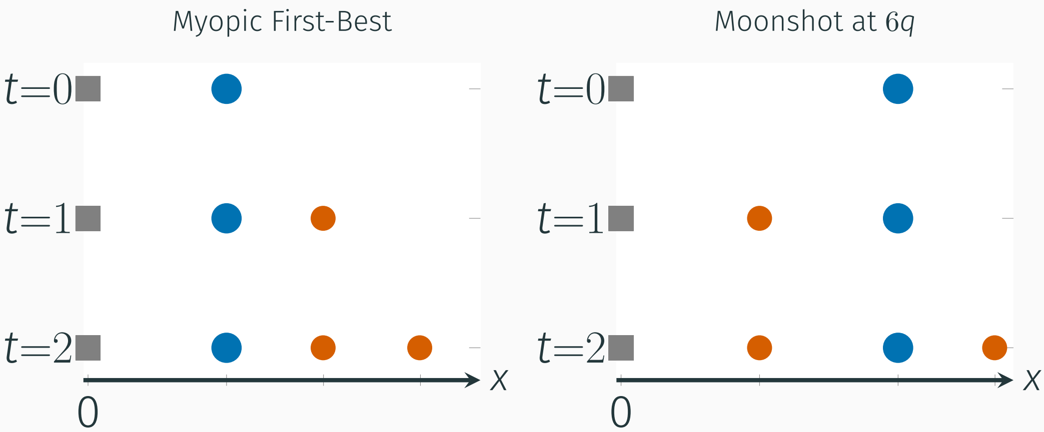
- $\mathcal{F}_0 = (0, y(0))$
- at $t = 0$, a designer can disclose 1 answer
- from $t = 1$ onward we play the dynamic game
- society receives flow payoff also in $t = 0$

What is best for society?

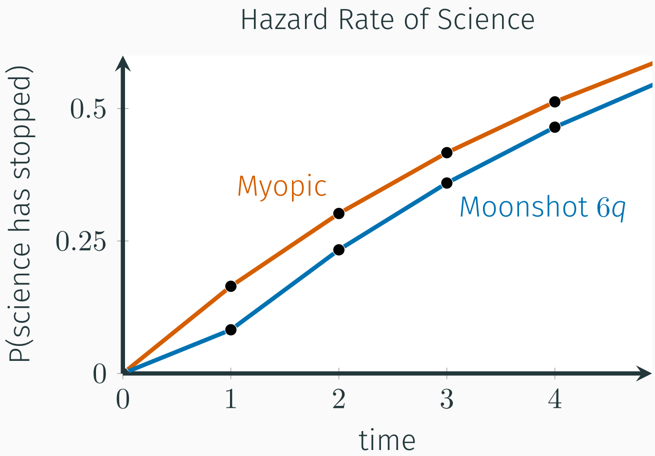
myopic? maximize period 0 payoff by disclosing $d = 3q$

hyperopic? guide $t > 0$ researchers and disclose a moonshot (d>3q)

Example $\eta = 1/8$ (Novelty)



Example: $\eta = 1/8$ (Output)



$\Rightarrow$ Moonshot provides more valuable landscape at higher probability!

When are Moonshots Optimal?

Are moonshots always beneficial? No.

As $\eta \to 0$: area length near optimal anyways, prob. of discovery high

As $\eta \to \infty$: very likely no discovery beyond moonshot

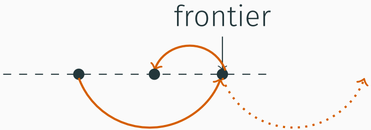
Proposition

There is a range $(\underline{\eta}, \overline{\eta}) \subset [0, \infty)$ such that for $\eta \in (\underline{\eta}, \overline{\eta})$ a moonshot is optimal at $t = 0$ provided δ is larger than a critical discount factor $\underline{\delta}(\eta) < 1$. The moonshot ignites a research cycle.

Moonshots: Main Take-Away

First Best and Laissez-fair:
Ladder structure(w/ different stepsize)

Moonshot:
Research Cycle



$\Rightarrow$ Iff patience high and cost intermediate cycles $\succ$ ladder.

funding moonshots

Repeated Moonshots

Suppose disclosure opportunities arise randomly with iid probability $\lambda > 0$ each period

Suppose further they are rare $\lambda \to 0$.

What's the optimal disclosure strategy if an opportunity arises?

Result

Question: Should we have inter-linked research cycles?

Answer: No.

Question: Suppose there is a *knowledge blockade* due to repetitive failures to advance. Should we remove the blockade or change the direction of science?

Answer: For short cycles, yes.

Analytically: for "promising environments" ($\delta\rho^{\infty} > 1/2$).

Numerically: always.

Conclusion

Conclusion

Three contributions

1. Flexible framework to quantify the value of a discovery
    - depends on whether knowledge is expanded or deepened
    - depends on the degree of novelty
    - highest for deepening on areas of intermediate length
2. Characterize researcher's optimal decision
    - microfounded search process to determine optimal research effort
    - novelty and output endogenously linked
    - simple friction that causes too narrow too risk-averse research
3. Fostering the Evolution of Knowledge
    - guide future research via moonshots
    - break ladder structure if cost intermediate and patience high
    - induce: research cycles

Literature

- **Economics of science:**
  Aghion, Dewatripont and Stein (2008), Bramoullé and Saint-Paul (2010), Rzhetsky et al. (2015), Fortunato et al. (2018), Iaria, Schwarz and Waldinger (2018), Liang and Mu (2020), ...

- **Discovering a Brownian path:**
  Rob and Jovanovic (1990), Callander (2011), Garfagnini and Strulovici (2016), Callander and Clark (2017), Prendergast (2019), Bardhi (2020), Bardhi and Bobkova (2021), Callander, Lambert and Matouschek (2022),...

Appendix

d vs. ρ: Substitutes and Complements

Output & Novelty: Substitutes or Complements?

Proposition

Suppose $\eta > 0$.

1. *When the researcher expands knowledge, distance, d, and probability of discovery, ρ, are substitutes.*
2. *When the researcher deepens knowledge, d and ρ are*
    - *independent if $X \leqslant 4q$,*
    - *complements if $X \in (4q, \frac{5}{2-\sqrt{3/2}}q)$,*
    - *substitutes for $d \in (0, \hat{d}(X))$ and complements for $d \in (\hat{d}(X), \frac{X}{2})$ if $X \in (\frac{5}{2-\sqrt{3/2}}q, 8q)$,*
    - *substitutes if $X > 8q$.*

Why Substitutes and Complements?

A ceteris paribus increase in novelty affects both

and back

- the marginal benefit of ρ, $V(d;X)$, and
- the marginal cost of ρ, $\frac{\mathrm{d}}{\mathrm{d}\rho}\left(erf^{-1}(\rho)\right)^2\sigma^2(d;X)$.

Success probability and novelty are complements if

$$\frac{\mathrm{d}}{\mathrm{d}d}\left(\frac{V(d;X)}{\sigma^2(d;X)}\right) > 0 \iff \frac{V_d(d;X)}{V(d;X)} > \frac{\sigma_d^2(d;X)}{\sigma^2(d;X)}.$$

$\frac{\sigma_d^2(d;X)}{\sigma^2(d;X)}$ is increasing and concave in X.
For $X < 4q$, $V(d;X) \propto \sigma^2(d;X)$ implying that d and ρ are independent.

Why Substitutes and Complements?

A ceteris paribus increase in novelty affects both

and back

- the marginal benefit of ρ, $V(d; X)$, and
- the marginal cost of ρ, $\frac{\mathrm{d}}{\mathrm{d}\rho}\left(erf^{-1}(\rho)\right)^2 \sigma^2(d; X)$.

Success probability and novelty are complements if

$$\frac{\mathrm{d}}{\mathrm{d}d}\left(\frac{V(d; X)}{\sigma^2(d; X)}\right) > 0 \Longleftrightarrow \frac{V_d(d; X)}{V(d; X)} > \frac{\sigma_d^2(d; X)}{\sigma^2(d; X)}.$$

When X just exceeds $4q$, the increase in $\frac{V_d(d;X)}{V(d;X)}$ accelerates as questions addressed proactively that were not before. d and ρ are complements.
As X increases, $\frac{\sigma_d^2(d;X)}{\sigma^2(d;X)}$ dominates for small d where $\frac{\sigma_d^2(d;X)}{\sigma^2(d;X)}$ is highest implying that d and ρ are substitutes.

Why Substitutes and Complements?

A ceteris paribus increase in novelty affects both

and back

- the marginal benefit of ρ, $V(d;X)$, and
- the marginal cost of ρ, $\frac{\mathrm{d}}{\mathrm{d}\rho}\left(erf^{-1}(\rho)\right)^2\sigma^2(d;X)$.

Success probability and novelty are complements if

$$\frac{\mathrm{d}}{\mathrm{d}d}\left(\frac{V(d;X)}{\sigma^2(d;X)}\right) > 0 \iff \frac{V_d(d;X)}{V(d;X)} > \frac{\sigma_d^2(d;X)}{\sigma^2(d;X)}.$$

As $d \to X/2$, the marginal cost effect $\sigma_d^2 \to 0$ implying that if $V_d(d;x) > 0$ d and ρ are complements.
Whenever d is such that $V_d(d;X) < 0$, d and ρ are substitutes.

Funding

Funding Research

So far: Moonshots come at no cost $\Rightarrow$ Augment model by an initial funding stage.

Assume a funder with budget K has two instruments with relative price κ:

1. Cost reductions: lowering a researcher's cost by h, $\eta = \eta_0 - h$.
2. Prizes: awarding a prize ζ with probability $\min\{\frac{\sigma^2(d;\mathcal{F}_k)}{s}, 1\}$ where $s > 3q$.

Researcher's new problem

$$\max_{d,\rho} \rho\Big(V(d;\infty) + \frac{\sigma^2(d;\infty)}{s}\zeta\Big) - \eta(erf^{-1}(\rho))^2\sigma^2(d;\infty).$$

Feasible Set

Proposition

Under some regularity conditions (see paper) the research possibility frontier is

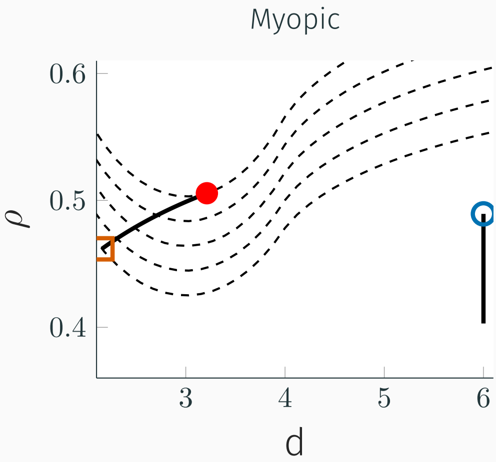
$$d(\rho; K) = 6q(K + s - \kappa\eta^0)\frac{\rho\tilde{c}_\rho(\rho) - \tilde{c}(\rho)}{2s\rho\tilde{c}_\rho(\rho) - s\tilde{c}(\rho) - \kappa\rho}.$$

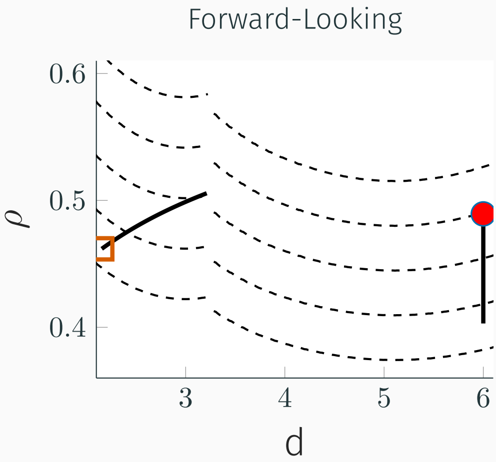
details in paper...

Myopic vs Forward-Looking Funding

Again: Consider $\mathcal{F}_1 = (0, y(0))$.

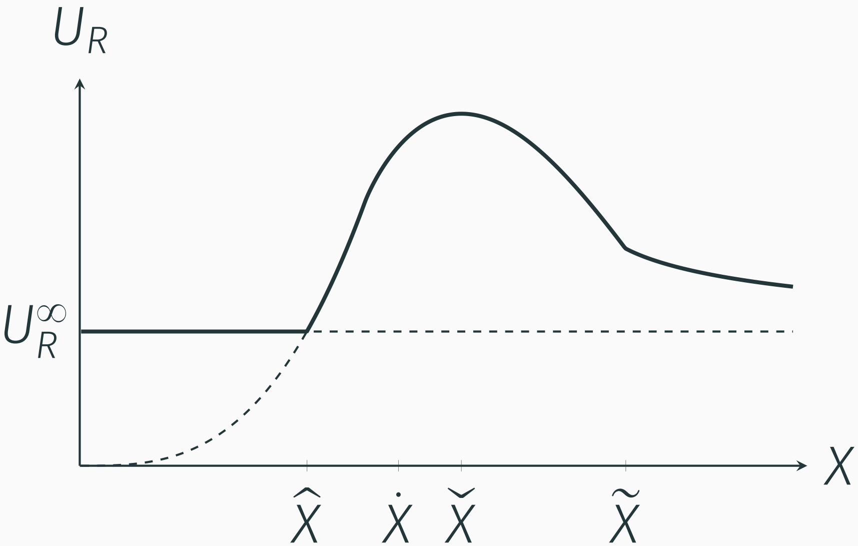
back



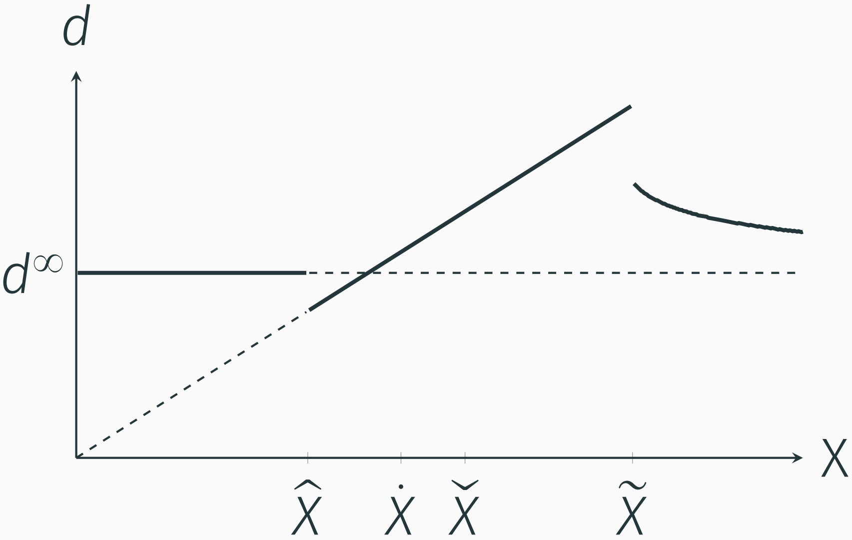


Graphs

Researcher's Value by Area Length



Novelty by Area Length



Output by Area Length

back