A Quest for Knowledge

Christoph Carnehl Johannes Schneider

February 2022

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 implications for the evolution of knowledge?
- How can funding institutions affect the researchers' actions?

Framework

We propose a microfounded model of knowledge and research in which:

1. Existing knowledge determines benefits and cost of research.
2. Successful research improves conjectures about similar questions.
3. Researchers are free to choose which questions to study and to what extent.

Society values knowledge as it informs complex decision making.

We conceptualize research as

- the selection of one out of many potential questions and
- the costly search for its answer.

Contribution

Our framework endogenously

- links the novelty of a research question to the probability that its answer is discovered
- determines when pushing the frontier is more valuable than bridging gaps between known results

Helps to address:

- **Evolution of knowledge:** dynamic externality of knowledge creation.
  $\rightarrow$ Short-run suboptimal novelty may improve the evolution of knowledge.
- **Science funding:** which choices can a budget-constrained funder implement?
  $\rightarrow$ Derive implementable set of output and novelty.

Literature

- **Economics of science:**
  Aghion, Dewatripont and Stein (2008), Bramoullé and Saint-Paul (2010), Foster, Rzhetsky and Evans (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), ...

Agenda

1. Model
2. Benefits of Discovery
3. Researcher's Choices
4. Funding Moonshots

Model

Model Basics

2 players, Researcher and Decision Maker

1. R observes initial knowledge.
2. R selects question and research intensity.
3. If R obtains a discovery, knowledge is augmented by it.
4. DM observes current knowledge and addresses a continuum of 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: $G_x(Y|\mathcal{F}_k)$.

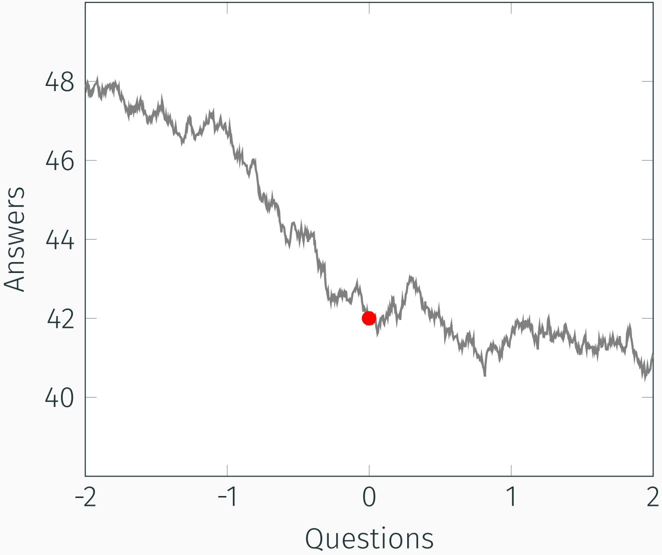
$\Rightarrow$ Brownian path determines answers: $Y(x|\mathcal{F}_k) \sim \mathcal{N}\left(\mu_x(Y|\mathcal{F}_k), \sigma_x^2(Y|\mathcal{F}_k)\right)$ with

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

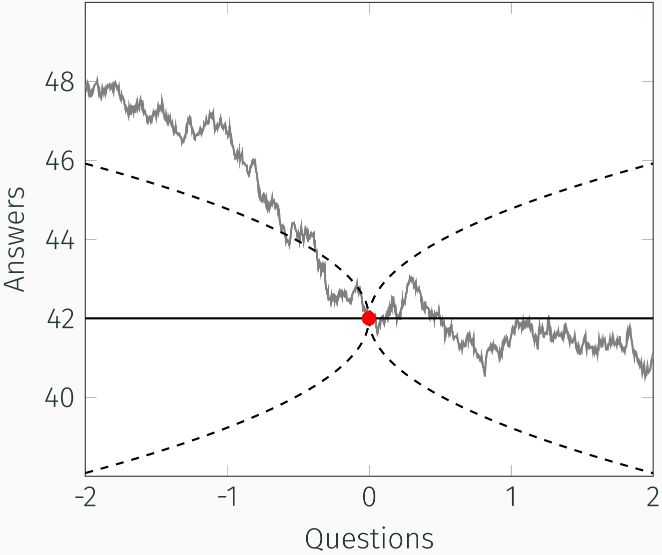
$$\sigma_x^2(Y|\mathcal{F}_k) = \begin{cases} x_1 - x & \text{if } x < x_1 \\ \frac{(x_{i+1} - x)(x - x_i)}{X_i} & \text{if } x \in [x_i, x_{i+1}) \\ x - x_k & \text{if } x \geqslant x_k. \end{cases}$$

Model of Knowledge - Graphically

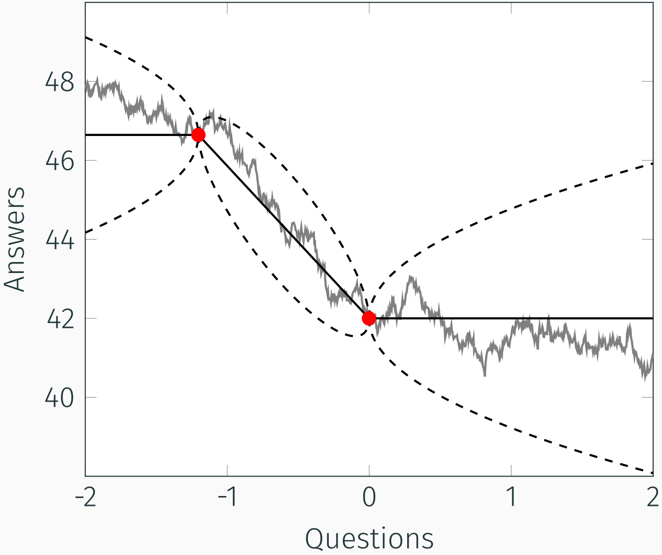
Truth and Knowledge



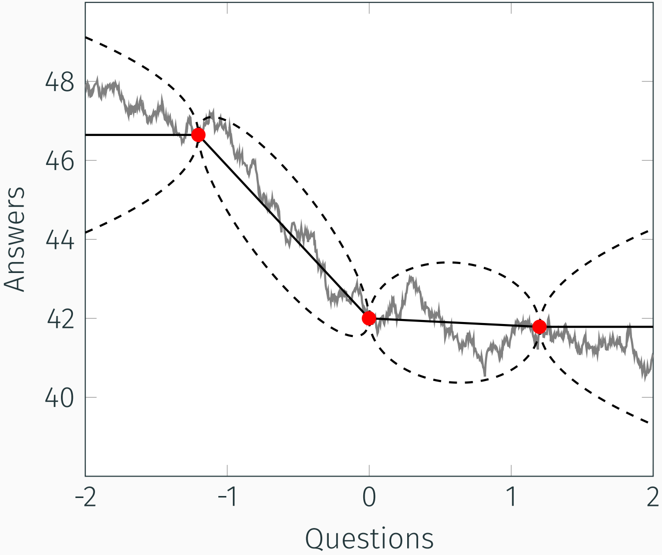
Conjectures



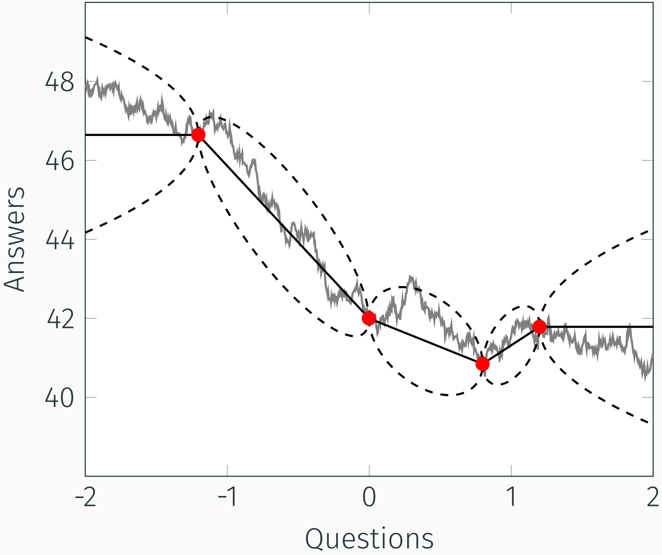
Expanding Knowledge



...to the Other Side



Deepening Knowledge



Society as Decision Maker

Decision Making

Society observes $\mathcal{F}_k$ 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.

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

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

Society's value of knowledge is

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

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 the improvement in society's 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

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

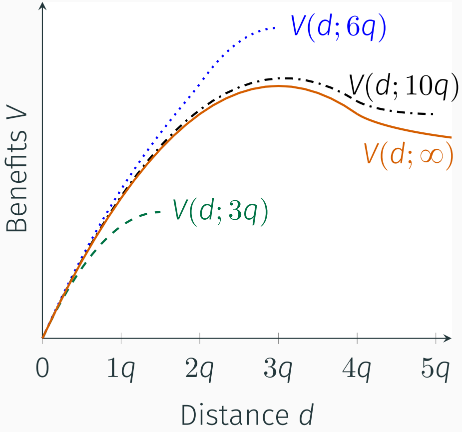
Benefit of discovery $V(d; X)$ determined by the question's distance to existing knowledge d and the length of the research area X.

Benefit of Discovery - Characterization

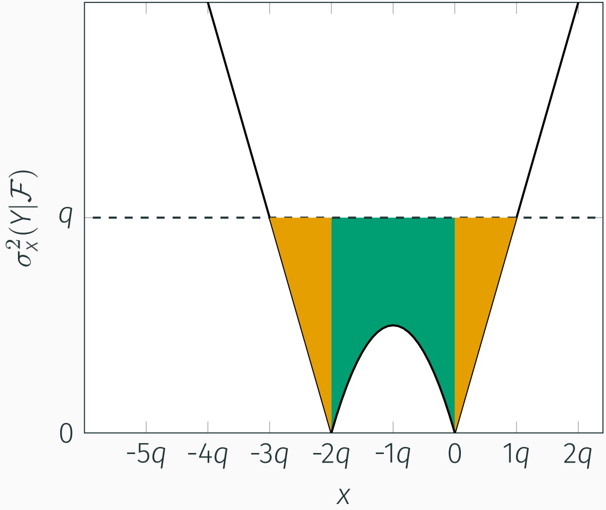
Proposition

Consider a discovery $(x, y(x))$ in a research area of length X with distance to existing knowledge d. The benefit of the discovery is

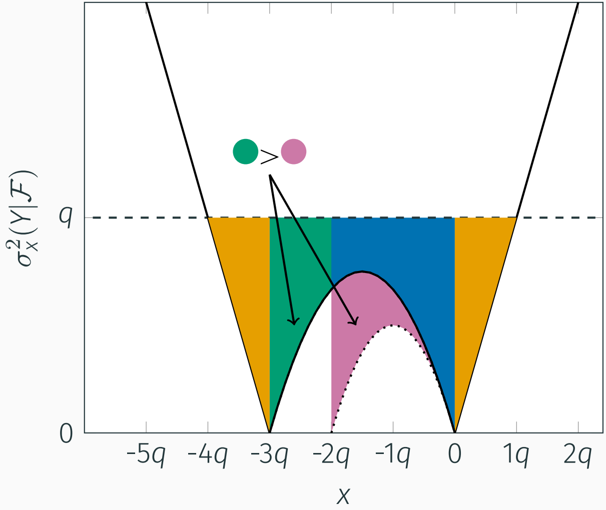
$$V(d;X) = \frac{1}{6q}\Big(2X\sigma^2(d;X) \\ + \mathbf{1}_{d>4q}\sqrt{d}(d-4q)^{3/2} \\ + \mathbf{1}_{X-d>4q}\sqrt{X-d}\,(X-d-4q)^{3/2} \\ - \mathbf{1}_{X>4q}\sqrt{X}(X-4q)^{3/2}\Big).$$



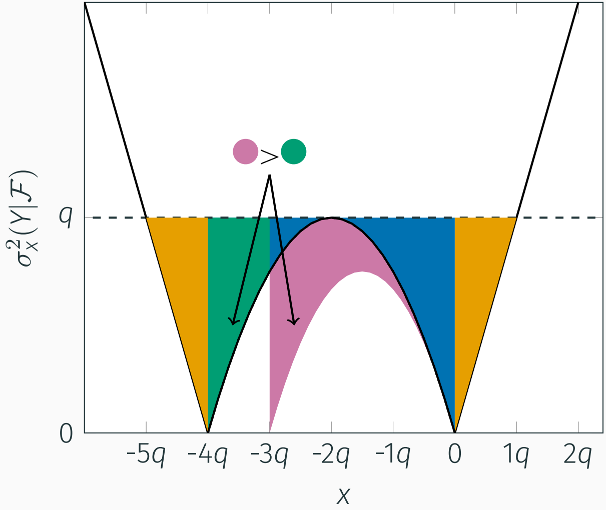
Benefit of Expanding Knowledge



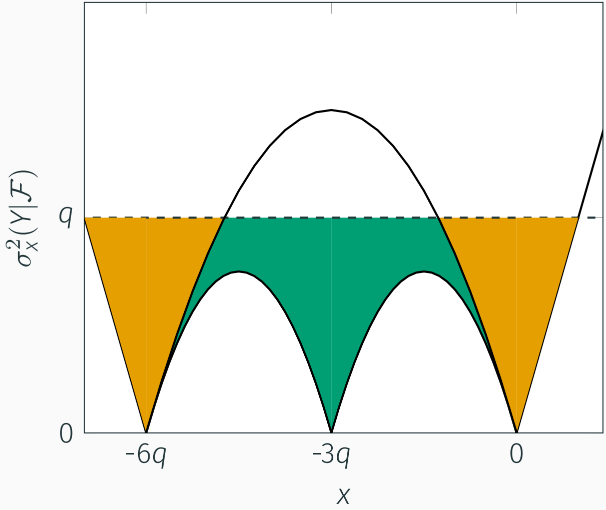
Benefit of Expanding Knowledge



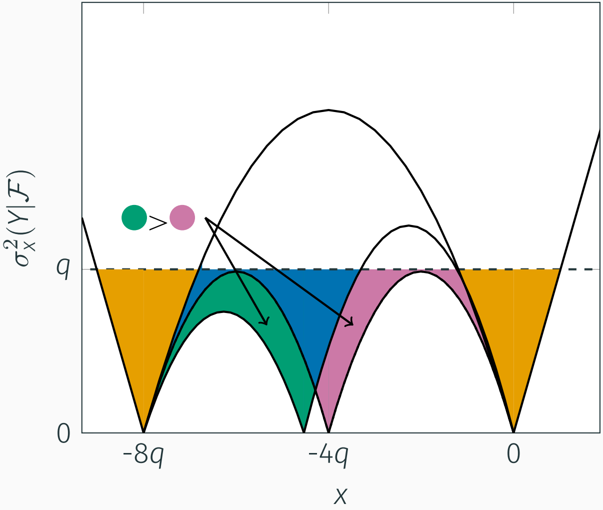
Benefit of Expanding Knowledge



Benefit of 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:

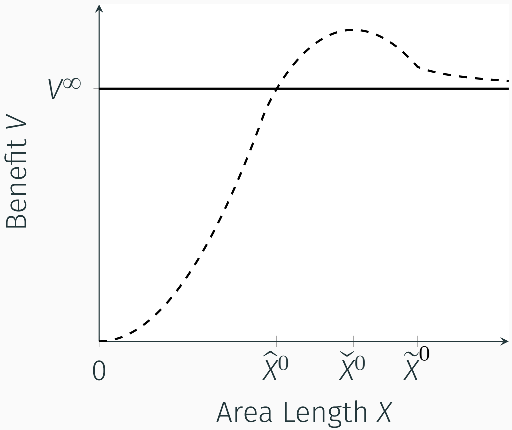
- *If* $X = \infty$, $d^0(\infty) = 3q$.
- *If* $X \leqslant \widetilde{X}^0 \in (6q, 8q)$, $d^0(X) = X/2$.
- *If* $X \in (\widetilde{X}^0, \infty)$, $d^0(X) \in (3q, X/2)$.
- $d^0(X)$ *is increasing in* X *for* $X < \widetilde{X}^0$ *and decreasing for* $X > \widetilde{X}^0$.

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: $c([a, b]) = \eta(b - a)^2$.

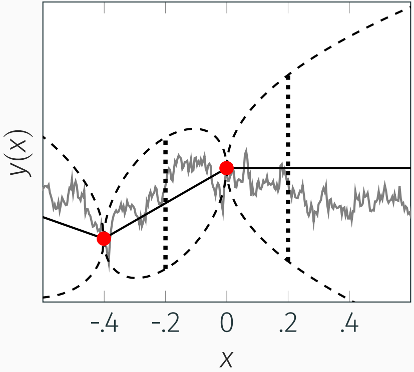
Lemma

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

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

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.

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

Optimal Choice: Distance, Novelty and Research Area

Proposition

Suppose $\eta > 0$. There is a set of cutoff values $\widehat{X} \leqslant \dot{X} \leqslant \check{X} \leqslant \widetilde{X} < 8q$ such that the following holds:

- *The researcher expands knowledge if and only if all available research areas are shorter than $\widehat{X}$.*
- *The researcher's payoffs, $U_R(X)$ are single peaked with a maximum at $\check{X}$.*
- *The optimal choices of distance, $d(X)$, and probability of discovery, $\rho(X)$, are non-monotone in X. The probability $\rho(X)$ has a maximum at $\dot{X}$, the distance $d(X)$ at $\widetilde{X}$.*

pictures

Researcher: Main Take-Away

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

Output choice ρ largest for intermediate areas.

Interaction: Novelty and output can substitute or complement each other.

- When expanding: substitutes
- When deepening
  - small areas: independent
  - medium-small areas: complements
  - medium-large areas: neither—substitutes (small distance) and complements (large distance)
  - large areas: substitutes

more details

Moonshots & Funding

Dynamic Game

Discrete time $t \in \{1, 2, ...\}$.

Players: long-lived DM & sequence of short-lived R

State: knowledge $\mathcal{F}_t$

Stage game: as before

DM aims to maximize $\max_{x,\rho} \ \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} v(\mathcal{F}_{t+1})\right]$.

Assumption

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

A few sanity checks

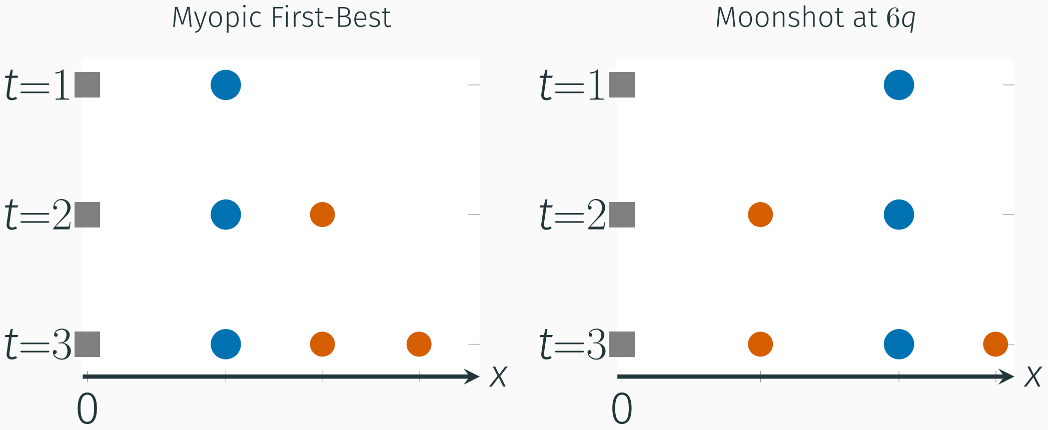
Corollary

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

- *first close gaps*
- *then aim at pushing the frontier step-by-step*
- *eventually end up failing to improve knowledge*

Moonshot ($\eta = 1/8$)

Thought experiment: $\mathcal{F}_1 = (0, y(0))$. Freely pick the first discovery.



Optimality of Moonshots

Are moonshots beneficial? Need to take into account effect on d and ρ.

- short-run loss $\rightarrow$ too large gap
- long-run gain $\rightarrow$ guidance for future researchers

Proposition

Suppose $\mathcal{F}_1 = (0, y(0))$. DM strictly prefers a moonshot in $t = 1$ for $\eta \in (\underline{\eta}, \overline{\eta})$ provided δ is larger than a critical discount factor $\underline{\delta}(\eta) < 1$.

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

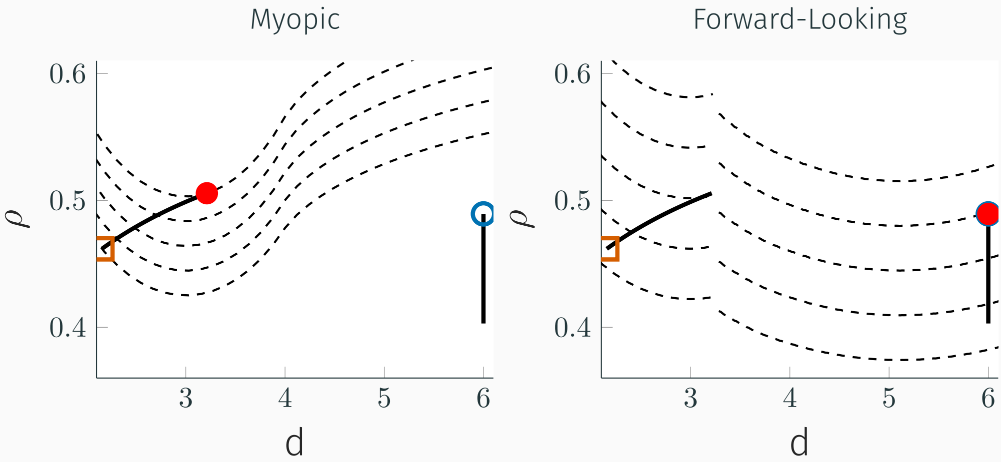
Under some regularity conditions (see paper) the research possibility frontier is

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

details in paper...

Myopic vs Forward-Looking Funding

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



Conclusion

Conclusion

Three contributions

1. 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
        - can be substitutes or complements for the researcher
        - when expanding always substitutes
3. Optimal funding of moonshots
    - moonshots guide future research
        - better knowledge
        - higher productivity
    - to fund moonshots, ex post rewards are needed
    - even myopic funder may combine ex post rewards with cost reductions to

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

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

$\frac{\sigma_d^2(d;X)}{\sigma^2(d;X)}$ 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

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

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, $\frac{\mathrm{d}}{\mathrm{d}\rho}\left(erf^{-1}(\rho)\right)^2 \sigma^2(d;X)$.

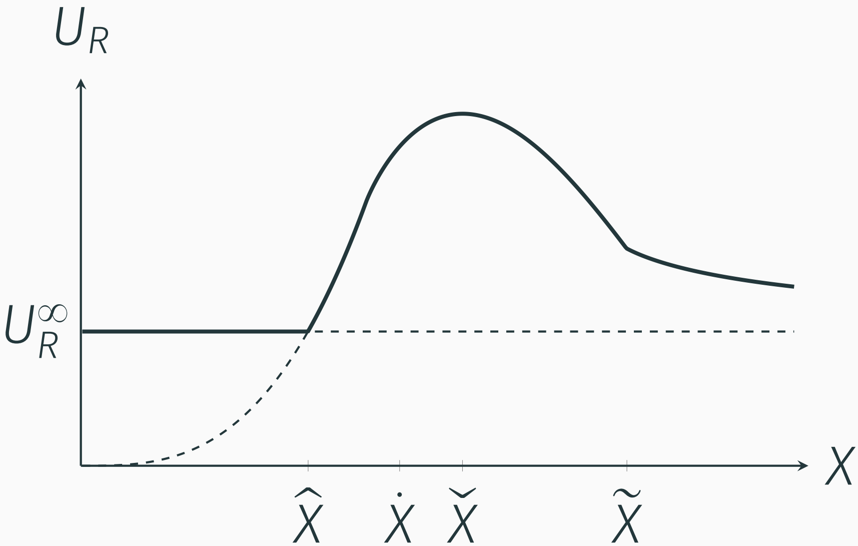
Success probability and novelty are complements if

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

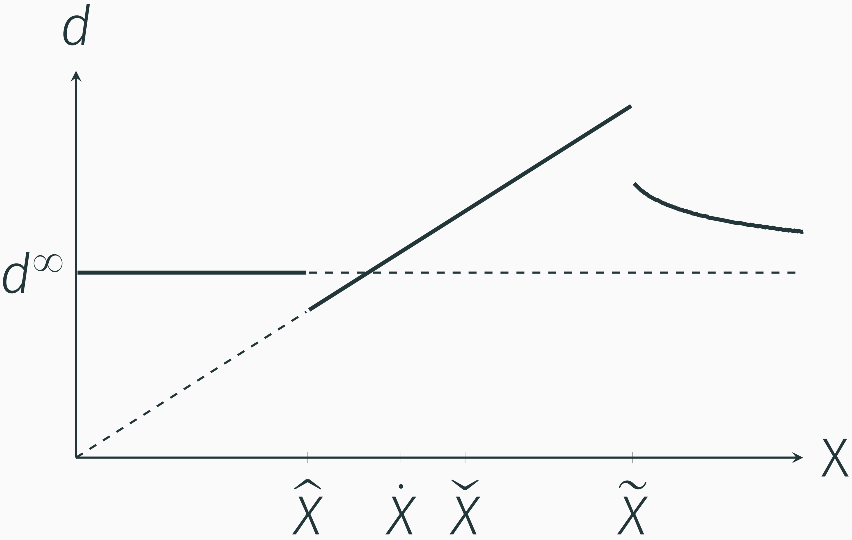
As $d \to X/2$, the marginal cost effect $\sigma_d^2 \to 0$ implying that if $V_d(d;x) > 0$ d and ρ are complements.
Whenever d is such that $V_d(d;X) < 0$, d and ρ are substitutes.

Graphs

Researcher’s Value by Area Length



Novelty by Area Length



Output by Area Length