A Quest for Knowledge

Johannes Schneider Christoph Wolf

May 2021

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 importance of research and 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 with three main features:

1. Existing knowledge determines benefit 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.

We conceptualize research as

- the selection of one out of many questions with correlated answers and
- the costly search for its answer building on existing knowledge.

Society values knowledge as it improves conjectures about optimal policies.

Contribution

Our framework endogenously links

- the novelty of a research question and
- the probability of discovering its answer.

Expanding the knowledge frontier is more desirable than deepening the existing knowledge only if the area between the frontiers is sufficiently well-understood.

Apply model to classical topics in the economics of science:

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

Literature

- **Science 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 of Knowledge
2. Benefit of Discovery
3. Cost of Research
4. Researcher's Choices
5. Application: Moonshots
6. Application: Science Funding

Model of Knowledge

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

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

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: $G_x(Y|\mathcal{F}_k)$.

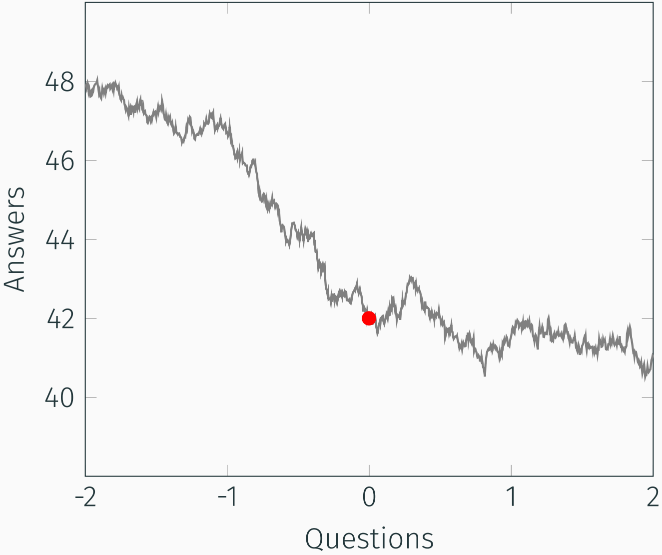
Brownian path determines answers: $Y(x) \sim \mathcal{N}\left(\mu_x(Y|\mathcal{F}_k), \sigma_x^2(Y|\mathcal{F}_k)\right)$ with

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

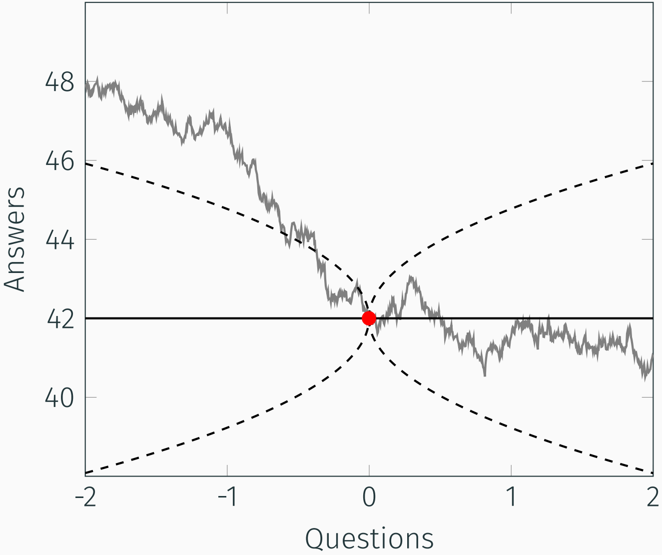
$$\sigma_x^2(Y|\mathcal{F}_k) = \begin{cases} x_1 - x & \text{if } x < x_1 \\ \frac{(x_{i+1}-x)(x-x_i)}{X_i} & \text{if } x \in [x_i, x_{i+1}) \\ x - x_k & \text{if } x \geqslant x_k. \end{cases}$$

Model of Knowledge - Graphically

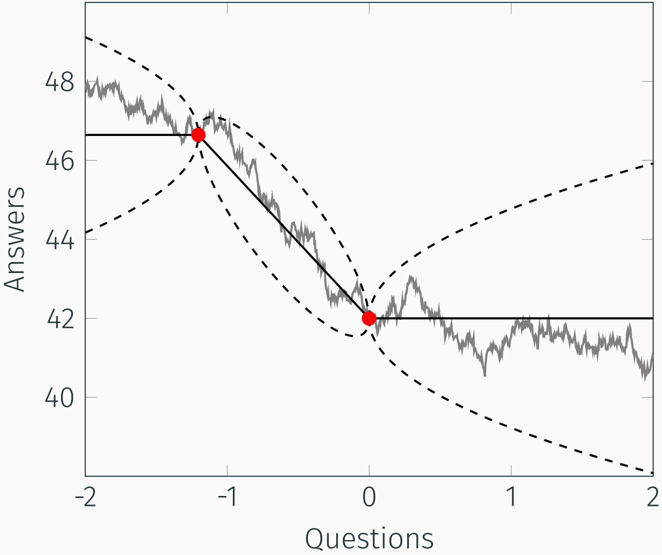
Truth and Knowledge



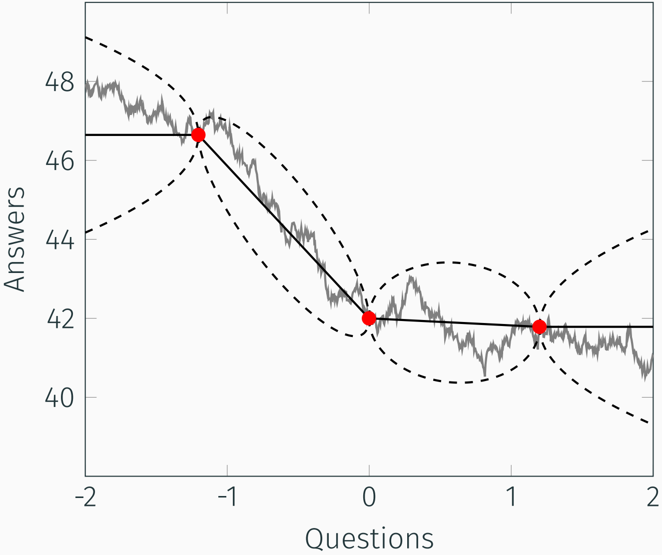
Conjectures



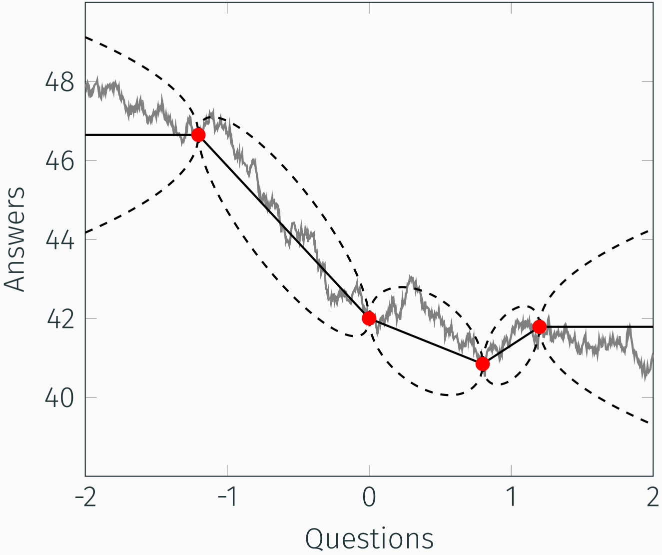
Expanding knowledge...



…on both sides



Deepening Knowledge



Society as Decision Maker

Decision Making

We represent society by a single decision maker that observes knowledge, $\mathcal{F}_k$. Society makes decisions on all questions, $x \in \mathbb{R}$, and can either

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

with per-question payoffs

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

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.

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

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

$$\mathcal{F}_{k+1} = \mathcal{F}_k \cup (x, y(x)).$$

The benefit of a discovery is the improvement in society's decision making

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

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

The problem simplifies 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 determined by the question's distance to existing knowledge d and the length of the research area X, $V(d; X)$.

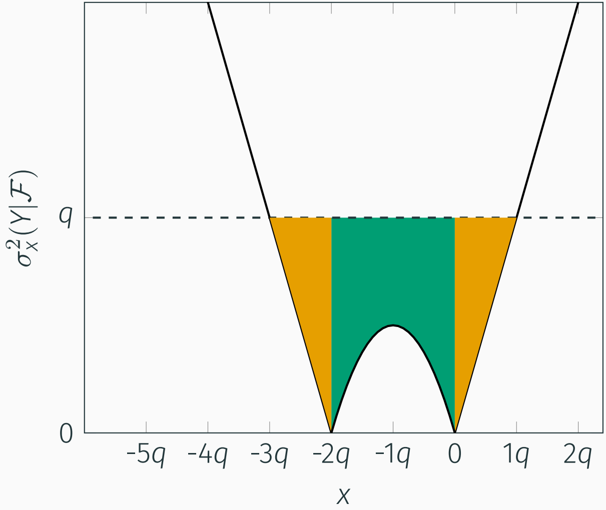
Benefit of Discovery - Characterization

Proposition

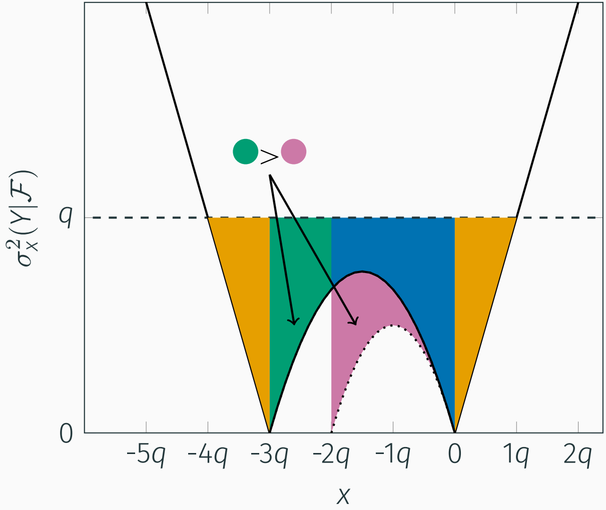
Consider a discovery $(x, y(x))$ in a research area of length X with distance to existing knowledge d. The benefit of the discovery is

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

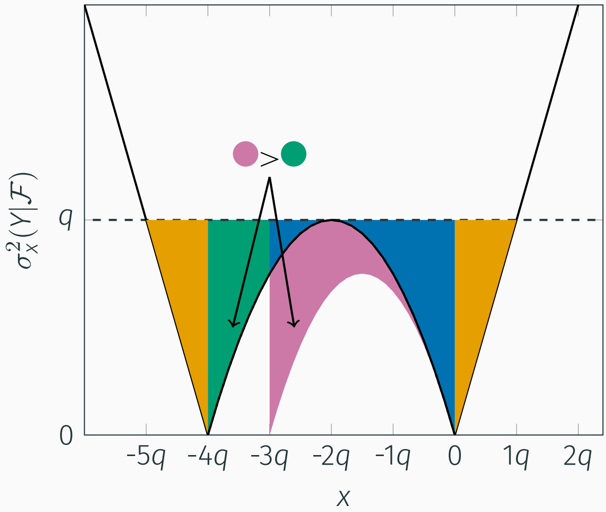
Benefit of Expanding Knowledge



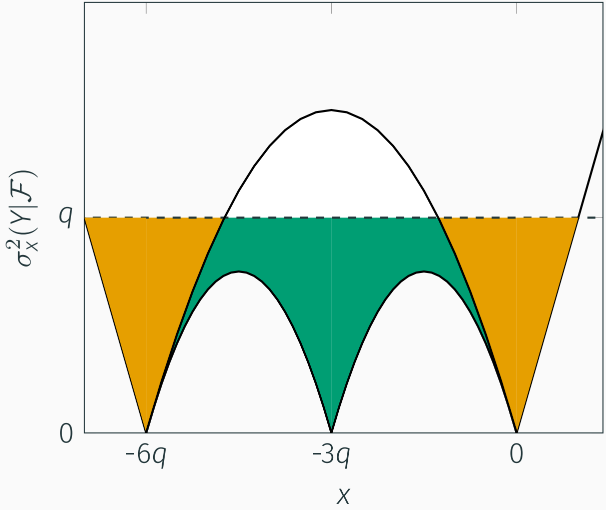
Benefit of Expanding Knowledge



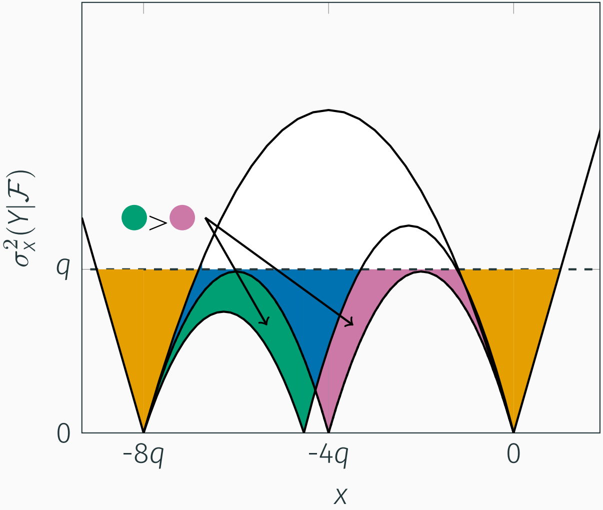
Benefit of Expanding Knowledge



Benefit of Deepening Knowledge



Deepening Knowledge



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 $X > \widetilde{X}^0$.*

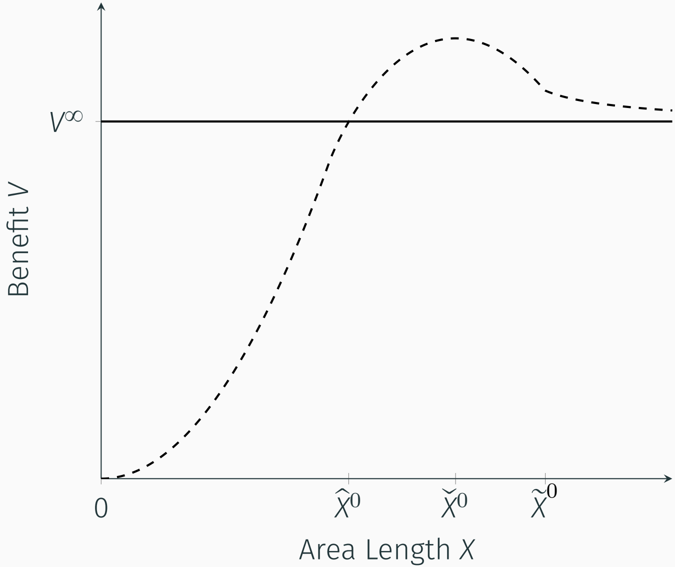
Properties of Benefit of Discovery

Corollary

There are two cutoff area lengths, $4q < \widehat{X}^0 < 6q < \check{X}^0 < 8q$, such that:

- *The maximum benefit of deepening knowledge in an area i is increasing in the area length if $X_i < \check{X}^0$; it is decreasing if $X_i > \check{X}^0$.*
- *The benefit of optimally expanding knowledge by $3q$ dominates the benefit of deepening knowledge in area i if and only if $X_i < \widehat{X}^0$.*

Benefit of Discovery by Area Length



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

Proposition

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 length

$$2^{3/2} erf^{-1}(\rho)\sigma(d; X)$$

and cost

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

Cost function is separable in probability ρ and precision of conjectures about $y(x)$.

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.

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

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

- 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

- 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

- 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

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

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.

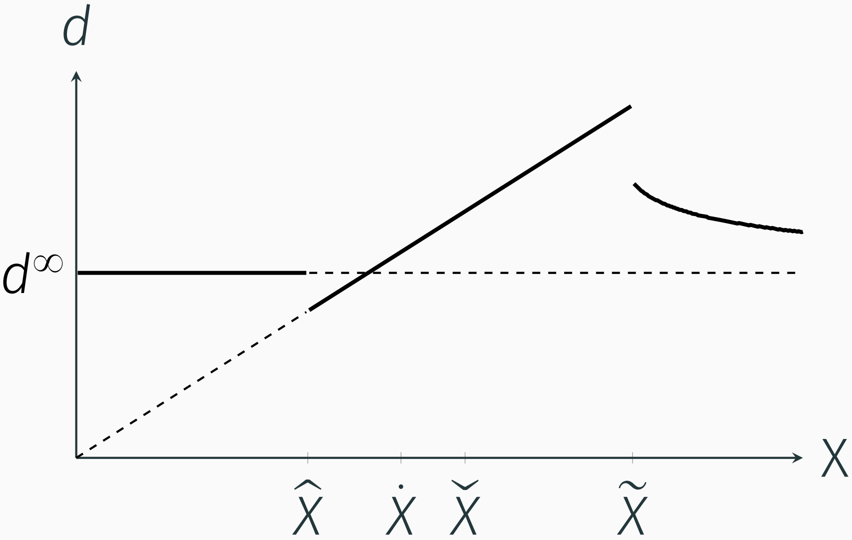
Optimal Choice: Distance, Novelty and Research Area

Proposition

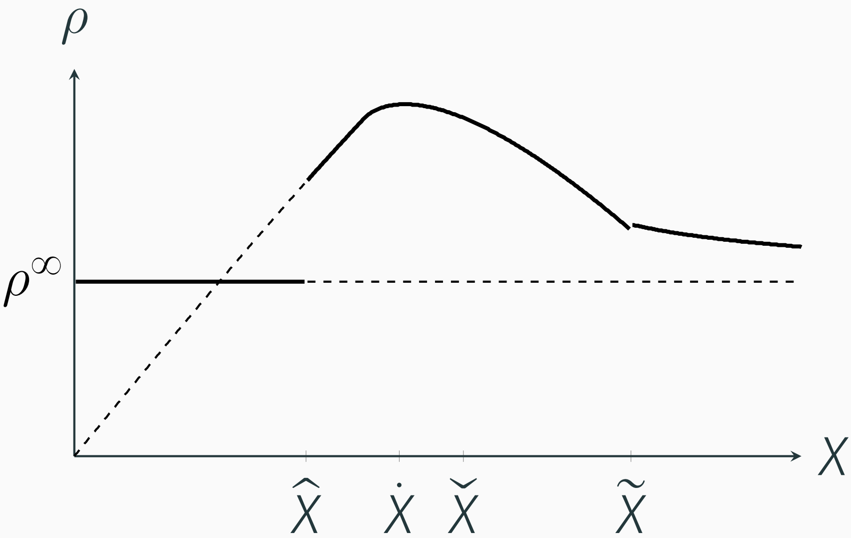
Suppose $\eta > 0$. There is a set of cutoff values $\widehat{X} \leqslant \dot{X} \leqslant \check{X} \leqslant \widetilde{X} < 8q$ such that the following holds:

- *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 $\widetilde{X}$.*

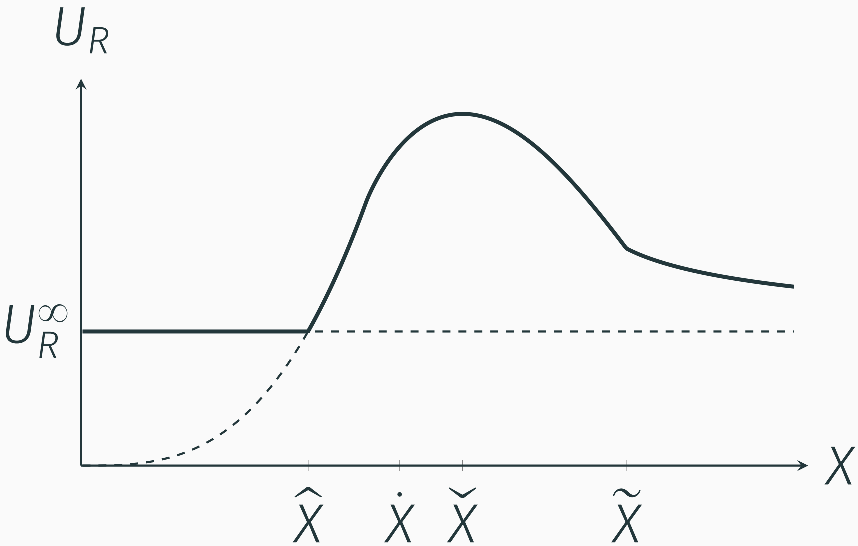
Novelty by Area Length



Output by Area Length



Researcher’s Value by Area Length



Takeaways: Researcher

The microfounded model of research provides insights for classical topics in the economics of science.

1. Output and novelty are endogenously linked via the cost of research and existing knowledge. They can be substitutes or complements.
   When expanding knowledge: more novelty $\rightarrow$ more risk.
2. Knowledge determines choice of research area, novelty and output.
   Research areas of intermediate length have high novelty and output.
3. When knowledge is generated by a sequence of short-lived researchers, there is a dynamic externality:
   Research today affects the benefit of discoveries and the success probabilities in the future.

Application: Moonshots

Incentivizing Moonshots

Is it beneficial to a long-lived society to incentivize very novel research; i.e., more novel than myopically optimal, $d > 3q$?

Consider the following variant of our model.

- Time is discrete, $t = 1, 2, ...$, with initial knowledge $\mathcal{F}_1 = \{0, y(0)\}$.
- Society discounts time by $\delta \in [0, 1)$.
- Sequence of short-lived researchers that observe $\mathcal{F}_t$ and decide on x and ρ. Assume symmetric strategies: Same $\mathcal{F}_t \Rightarrow$ same choice of x and ρ.
- If research is successful, knowledge updates to $\mathcal{F}_{t+1} = \mathcal{F}_t \cup (x, y(x))$. Otherwise $\mathcal{F}_{t+1} = \mathcal{F}_t$.

In period 1, society can costlessly pick x and ρ.

Optimality of Moonshots

Society maximizes

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

where choices in periods $t \geqslant 2$ are made by researchers individually.

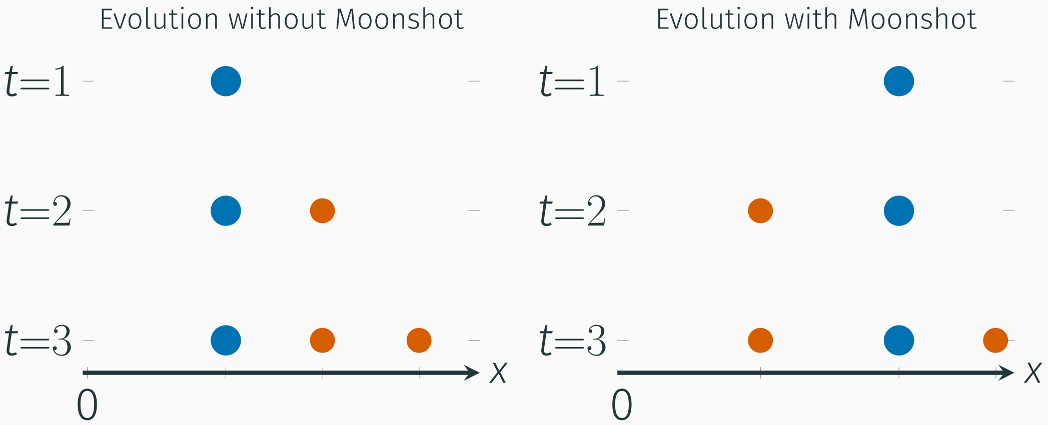
Proposition

There is a non-empty interval $(\underline{\eta}, \overline{\eta})$ such that the decision maker strictly prefers a moonshot in $t = 1$ for any $\eta \in (\underline{\eta}, \overline{\eta})$ provided δ is larger than a critical discount factor $\underline{\delta}(\eta) < 1$.

Low cost $\rightarrow$ small distortion $\rightarrow$ short-run losses dominate.

High cost $\rightarrow$ small benefit of moonshot $\rightarrow$ short-run losses dominate.

Moonshots and Evolution of Knowledge ($\eta = 1/8$)



Application: Science Funding

Science Funding

Under scientific freedom, researchers can freely choose their research questions.

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 $s > 3q$.

Which novelty-output combinations can a budget-constrained funder incentivize?

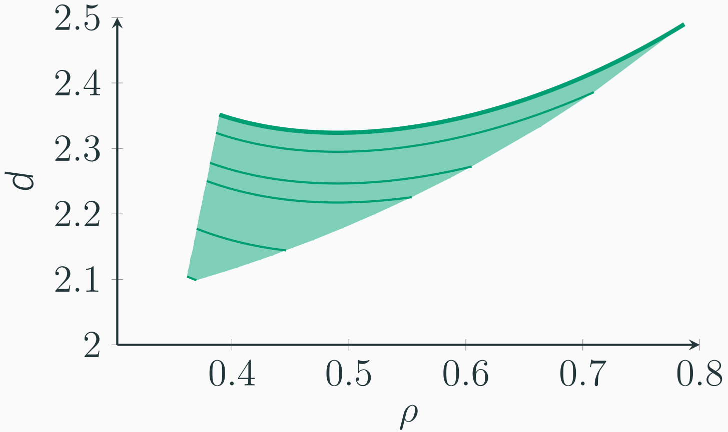
Proposition

The research-possibility frontier $d(\rho;\kappa,K)$ defined over $[\underline{\rho}(\kappa,K), \overline{\rho}(\kappa,K)]$

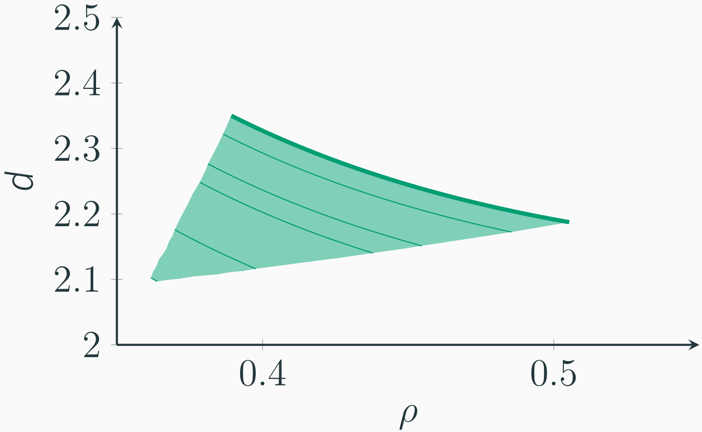
$$d(\rho;\kappa,K) = \min\{6q(K + s - \kappa\eta^0)\frac{\rho\tilde{c}_\rho(\rho) - \tilde{c}(\rho)}{2s\rho\tilde{c}_\rho(\rho) - s\tilde{c}(\rho) - \kappa\rho}, s\}.$$

To incentivize any $d > 3q$, the funder must award prizes for discoveries.

Research-Possibility Frontier: Substitutes and Complements



Research-Possibility Frontier: Substitutes and Complements



Maximizing Benefit of Discovery

Consider a funding institution that maximizes the static benefit of expanding knowledge:

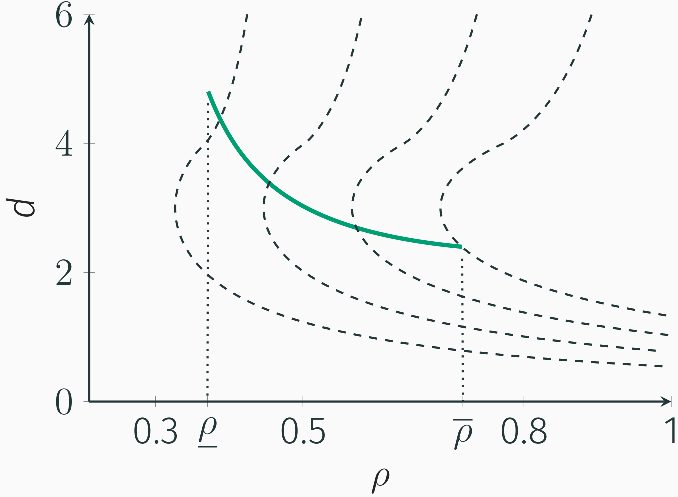
$$\max_{\zeta,\eta} \quad \rho V(d;\infty)$$
$$\text{s. t.} \quad \zeta + \kappa(\eta_0 - \eta) = K.$$

Proposition

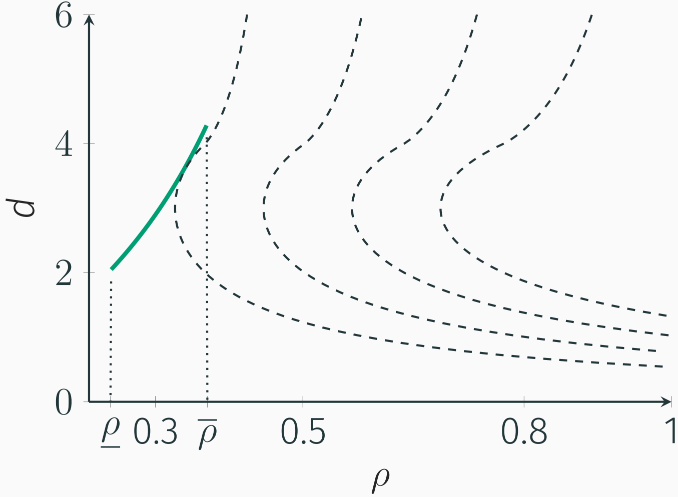
If output and novelty are substitutes from the funder's perspective, the optimal funding mix never incentivizes excessive novelty, $d > 3q$.

If output and novelty are complements from the funder's perspective, the optimal funding mix might incentivize excessive novelty, $d > 3q$.

Optimal Funding - Cost Reduction Cheap



Optimal Funding - Cost Reduction Expensive



Conclusion

Conclusion

We propose a model of knowledge built on

1. a large pool of questions,
2. knowledge informing conjectures about related questions,
3. society applying knowledge to choose policies.

We conceptualize research as the

1. free choice of research questions and
2. and the costly search for their answers.

Our model

- endogenously links novelty and research output and
- highlights the importance of existing knowledge for research and knowledge accumulation.

Applications of Model

Model is tractable and widely applicable in the economics of science.

1. Evolution of science
   - Dynamic externality $\rightarrow$ suboptimal knowledge accumulation.
   - Moonshots can improve evolution of knowledge.
2. Optimal research incentives
   - Novelty and output can be complements for funding institutions.
   - Cost reductions alone cannot incentivize moonshots.
3. Consequences of null results (work in progress)
4. Innovation, competition, ...