A Quest for Knowledge

Johannes Schneider (uc3m) Christoph Wolf (Bocconi)

September 2020

Motivation

[Solutions] come through asking the right questions, because the answers pre-exists. It is the questions that we must define and discover. [...] You don't invent the answer—you reveal the answer.

Jonas Salk

What are 'the right question'?

For the researcher: What gives me recognition?

productivity: what is easy to answer?

novelty: how far away is it from what is known?

$\Rightarrow$ potential trade-off: novelty v productivity

For a decision maker: What improves decision making?

productivity: only findings help to improve decision making

direct effect: the answer to question a solves the problem a

spill-over effects: can the answer to a teach us about problem b?

Funding & the Shoulders of Giants

Incentives driven by

Path dependency: "Standing on the shoulders of giants"

Science funding: ex-ante cost reductions vs ex-post rewards

This paper

Propose a natural model of sciences based on:

1. knowledge is informative for decision making,
2. knowing the answer to certain questions has an *externality* on the conjectures on related question, and
3. the set of questions available is infinite and the impact of answering a question on conjectures about the answers to another question depends on how close the two questions are.

Derive (in order of appearance):

1. an endogenous value-of-research function
2. an endogenous cost-of-research function
3. a characterization of the researcher's decision
4. a framework to study changes in the funding architecture

Outline

Model

Model Ingredients

Questions and Answers: question $x \in \mathbb{R}$ with associated answer $y(x) \in \mathbb{R}$

Truth: the collection of answer to all questions

Knowledge: the collection of answers to which the question is known.
$\mathcal{F}_k = \{(x_i, y(x_i))\}_{i=1}^{k}$ w/ $x_i < x_{i+1}$.

Truth-Generating Process: Brownian motion $Y(x)$ over the real line

Some notation:

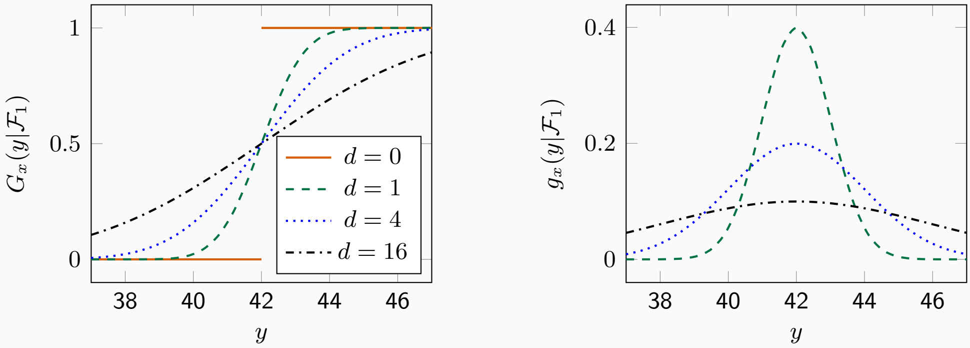
$d(x)$: (Euclidean) distance of question x to the nearest known point,
$d(x) = \min_{x_i \in \mathcal{F}_k} |x - x_i|$.

X_i : Distance between two known points $|x_{i+1} - x_i|$

$\mathcal{X}_k$: Collection of all X_i in $\mathcal{F}_k$

First Implication: Conjectures

Knowledge $\mathcal{F}_k$ allows us to form conjectures, $G_x(y|\mathcal{F}_k)$ about the answer to *all* other question. All conjectures are normally distributed.



Conjectures Properties

Property 1. Given knowledge $\mathcal{F}_k$, the answer to question x has expectations

$$\mu_{\mathcal{F}_k}(x) = \begin{cases} y(x_1) & \forall\ x < x_1 \\ y(x_i) + \frac{x - x_i}{X_i}(y(x_{i+1}) - y(x_i)) & \forall\ x \in [x_i, x_{i+1}] \\ y(x_k) & \forall\ x > x_k. \end{cases}$$

Property 2. Given knowledge $\mathcal{F}_k$, the answer to question x has variance

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

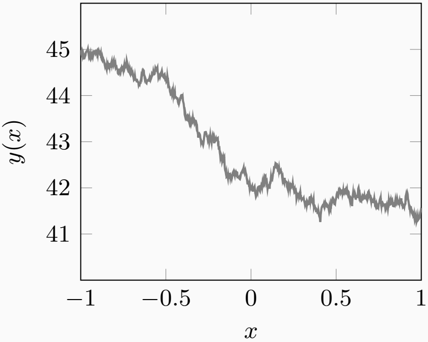
Decision Making

If a decision maker faces question x, she observes $\mathcal{F}_k$ and selects a response $a \in \mathbb{R} \cup \varnothing$.
The decision maker's payoff is

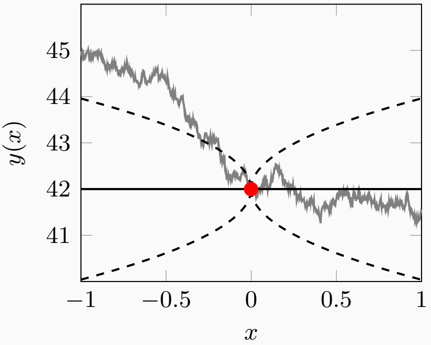
$$u(a;x) = \begin{cases} -(a - y(x))^2 & \text{if } a \neq \varnothing \\ -q & \text{if } a = \varnothing. \end{cases}$$

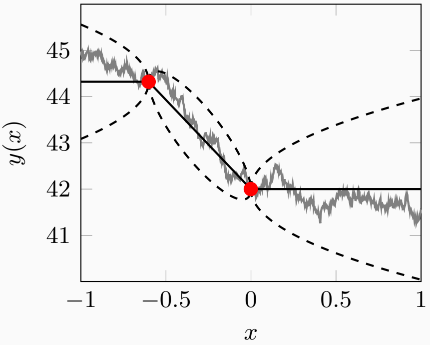
Example

The Color of the Truth is Grey

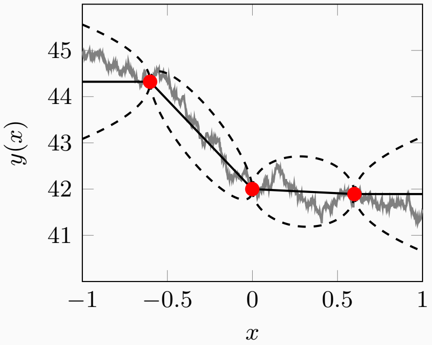


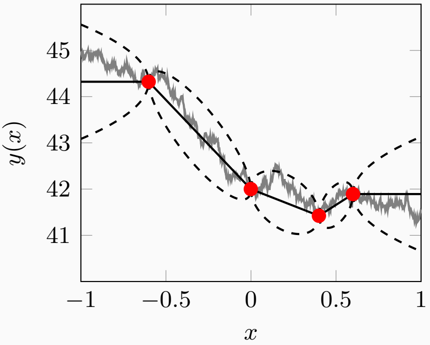
The Answer to the Questions on Life, the Universe, and Everything





Bridging the Gap





The Benefits of Research

Benefits of Research

Benefits from:

- new answers
- improved conjectures

Value of knowing $\mathcal{F}_k$ (from the DM's optimization problem)

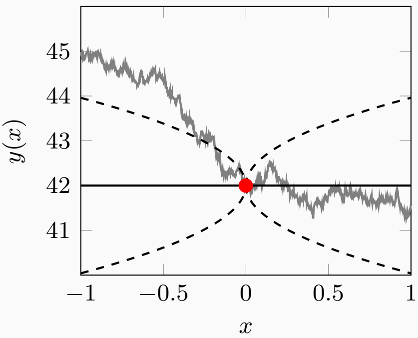
$$v\left(\mathcal{F}_k\right) = \int \max\left\{\frac{q - \sigma^2_{\mathcal{F}_k}(x)}{q}, 0\right\} dx.$$

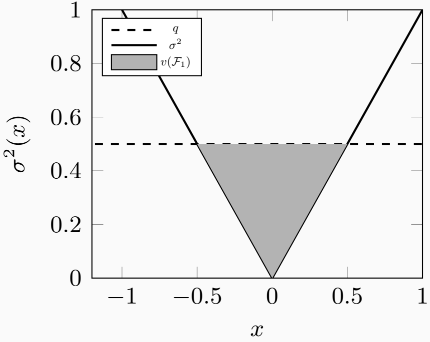
Value of moving from $\mathcal{F}_k$ to $\mathcal{F}_{k+1}$

$$V := v(\mathcal{F}_{k+1}) - v(\mathcal{F}_k)$$

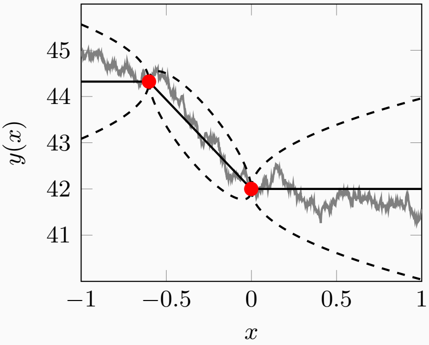
Two types of knowledge generation: expanding (on interval length $X = \infty$) or deepen knowledge (on interval $X < \infty$).

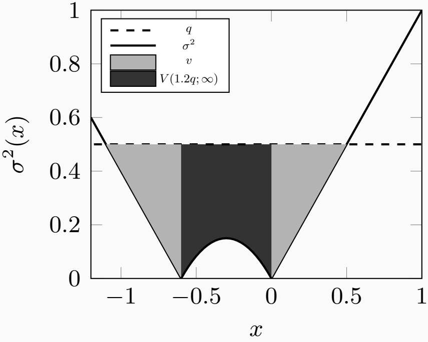
Example: The Value of Knowing of $\mathcal{F}_k$.



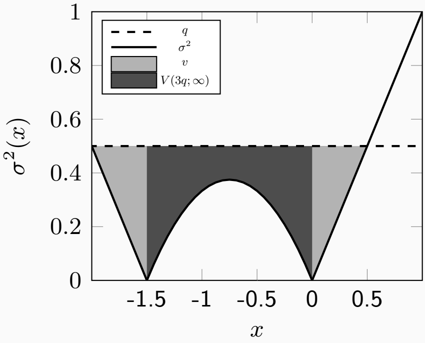


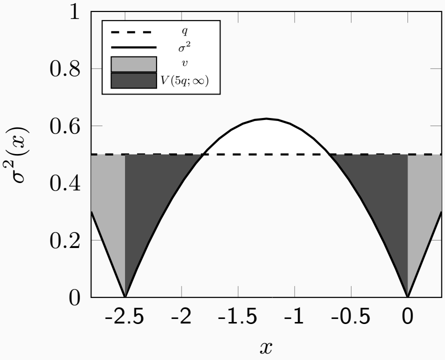
Example: The Benefits of Research





Example: Different choices in x





The Benefits of Research

Proposition

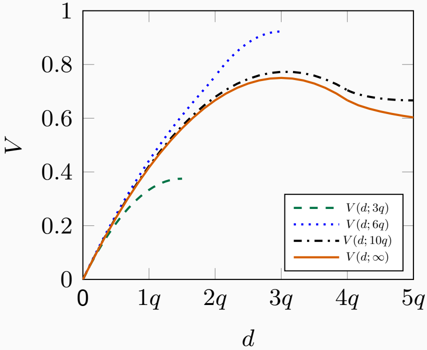
Suppose research obtains the answer to any question with distance $d \equiv d(x)$. The benefit of expanding is

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

that of deepening is

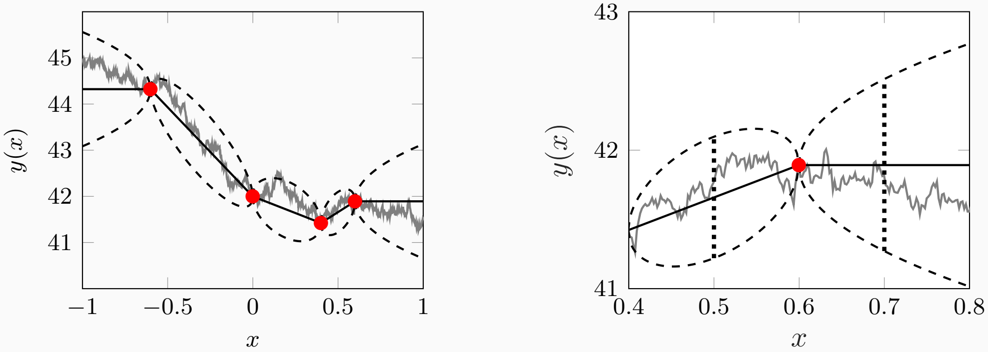
$$\begin{aligned} V(d;X_i) = \frac{1}{6q}\Big(2dX_i - 2d^2 &+ \mathbf{1}_{d>4q}\sqrt{d}(d-4q)^{3/2} \\ &- \mathbf{1}_{X_i>4q}\sqrt{X_i}(X_i-4q)^{3/2} \\ &+ \mathbf{1}_{X_i-d>4q}\sqrt{X_i-d}\,(X_i-d-4q)^{3/2}\Big). \end{aligned}$$

The Benefits of Research



Cost of Research

Sampling Intervan



Prediction Interval

suppose the cost of sampling interval $[a, b]$ are $(a-b)^2$

Definition

The prediction interval $\alpha(x, \rho)$ is the smallest interval $[a, b] \subseteq \mathbb{R}$ such that the answer to question x lies within $[a, b]$ with a probability of at least ρ.

Cost of Research

Lemma

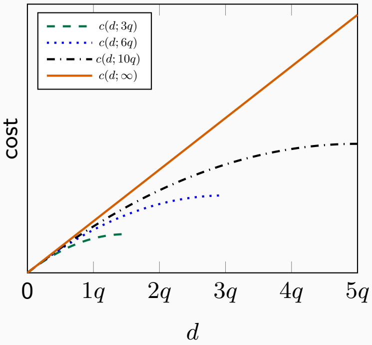
Suppose $\alpha(x, \rho)$ is the prediction interval for probability ρ and x is normally distributed with mean μ and standard deviation σ. Then any prediction interval has the following two features:

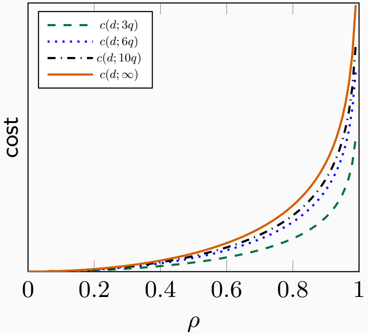
1. The length of the prediction interval is $2^{3/2} erf^{-1}(\rho)\sigma$,
2. The interval is centered around μ,

where erf^{-1} is the inverse of the Gaussian error function.

$$\Rightarrow c(\rho, d) = 8(erf^{-1}(\rho))^2\sigma^2(x; \mathcal{F}_k).$$

Cost: Comparative Statics





Researcher's Choice

Researcher's Problem

Two parameters

η: cost intensity

ζ: ex-post reward for solving a difficult question

The researcher's problem is

$$\max_{X \in \mathcal{X}_k \cup \infty} \max_{\substack{d \in [0, X/2], \\ \rho \in [0,1]}} \rho \left(V(d;X) + \zeta \sigma^2(d;X) \right) - \eta c(\rho, d; X).$$

For now: restrict attention to situations in which the maximum interval $\max \mathcal{X}_k < 4q$.

Researcher's Choice

Proposition

Fix the existing knowledge $\mathcal{F}_k$ and consider the researcher's choice (X^*, d^*, ρ^*). A *marginal decrease* in the cost parameter, η, implies that the length of the interval that contains the optimal question, X^*, weakly increases. Moreover, such a decrease in η

- (weakly) increases novelty, $\eta \downarrow \Rightarrow d^* \uparrow$, and strictly if $d^* \neq X^*/2$,
- (strictly) increases output on the intensive margin $\eta \downarrow \Rightarrow \rho^* \uparrow$ if X^* remains unchanged
- (strictly) decreases output on the extensive margin $\eta \downarrow \Rightarrow \rho^* \downarrow$ if X^* changes.

The effect of a *marginal increase* in the reward ζ is *qualitatively identical*.

Observations

- Ex-post rewards may lead to too much novelty (novelty islands)
- Ex-ante cost reductions always lead to a better outcome

Funding

Funding of Science

A funder takes the current incentive structure (ζ, η) as given and decides how to spend her money K to either

- reduce η (at cost κ)
- increase ζ (at cost 1)

Budget constraint:

$$K = z - \kappa * h$$

where $z :=$ increase in ex-post rewards
and $h :=$ reduction in ex-ante cost

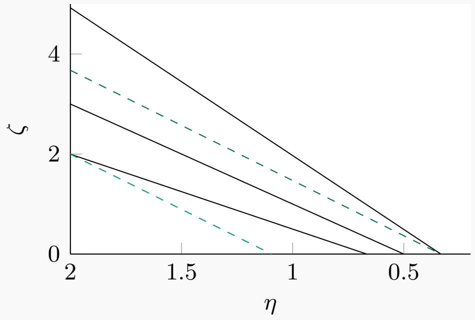
Exercise 1: Output maximization

Consider $\mathcal{F}_1$ and suppose the funder is mainly concerned about output maximization (i.e. increasing ρ). How should she spend her money?

Answer: It depends.

- If K is small $\Rightarrow$ all in ex-post rewards
- If K is large $\Rightarrow$ all in ex-ante cost reduction

Reason: Iso-ρ (derived from the FOC) curves are linear with increasing slope in ρ.



iso-ρ curves

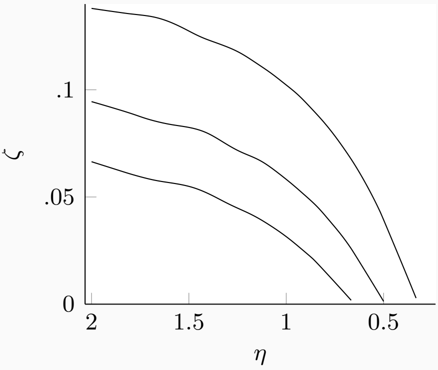
Exercise 2: Novelty Maximization

Consider $\mathcal{F}_1$ and suppose the funder is mainly concerned about novelty (i.e. increasing d). How should she spend her money?

Answer: It also depends. We know

- Corner solution.
- If all in ζ is optimal for ρ maximization, so is it for d maximization

Reason: Iso-d curves concave and smaller slope than Iso-ρ curves (for any level of ρ).



iso-d curves
(numerically approximated)

Exercise 3: General Objectives

Which (d, ρ) combinations does (K, κ) permit in general?

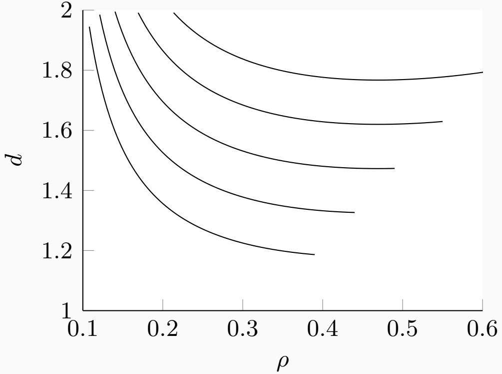


Figure: (ρ, d)-frontier for different budgets

What about deepening research?

- There is a cutoff $\tilde{X}$ such that the researcher
    - deepens research if $\exists X \in \mathcal{X} > \tilde{X}$
    - expands otherwise

That cutoff

- decreases in η
- increases in ζ

Moreover, if $\rho < 1/\kappa$ for all feasible ρ
$\Rightarrow \tilde{X} \uparrow$ if $\rho \uparrow$ (for a fixed budget)
$\Rightarrow$ we may see a reduction in ρ because the researcher switches from deepening to expanding

Conclusion

Summary and Road Ahead

Natural yet tractable framework to study:

- researcher's decision making
- spillovers in Knowledge
- changes in the funding architecture

More in the paper:

- heterogeneous researchers & the evolution of knowledge
- the value of publishing 0 results

Literature

Literature

Selection of Research Questions: Rzhetsky et al. (2015) and Fortunato et al. (2018)

Science Funding: Price (2019), Azoulay and Li (2020), and Price (2020)

Distance v Risk Taking: Iaria, Schwarz, and Waldinger (2018), Veugelers and Wang (2019), and Wilhelm (2019)

Researcher Incentives: Aghion, Dewatripont, and Stein (2008), Bramoullé and Saint-Paul (2010), Frankel and Kamenica (2019), and Frankel and Kasy (2020)

Methodology: Callander (2011), Callander and Clark (2017), and Bardhi (2019)