

Will That Idea Scale?

Ask ChatGPT!

John A. List
@Econ_4_Everyone

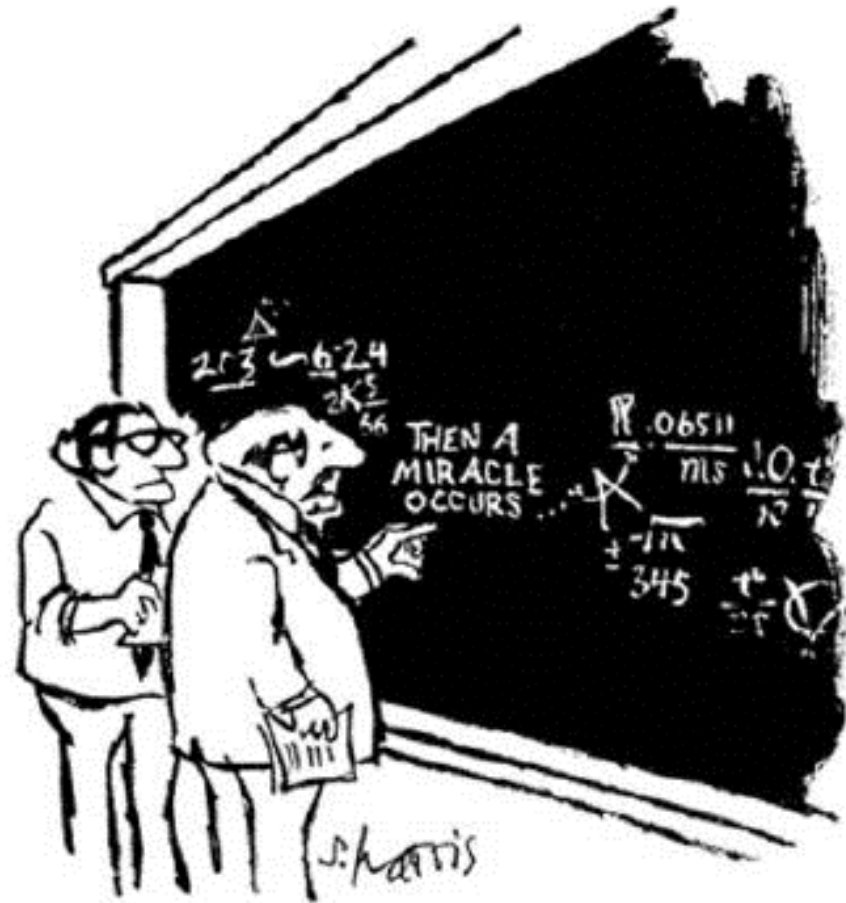
The Art of Scaling

The Economics of Scaling

John A. List
@Econ_4_Everyone

Science
The ~~Art~~ of Scaling

Scaling has been
a bit like this...



"I think you should be more explicit here in
step two."

Scaling



Disturbingly vague: bundles many different concepts that need to be understood before we can give advice to policymakers

External validity is about generalizing treatment effects

Scaling is about generalizing benefit-cost profiles

The Science of Using Science

- Al-Ubaydli, Omar, John A. List, and Dana Suskind. “What Can We Learn From Experiments? Understanding the Threats to the Scalability of Experimental Results,” *American Economic Review*, 107(5; P&P), 2017.
- Al-Ubaydli O., List, J.A., Danielle LoRe, and Dana Suskind. “Scaling for Economists: Lessons from the Medical Literature,” *Journal of Economic Perspectives*, 31(4), 2017.
- ❖ Al-Ubaydli, Omar, John A. List, and Dana Suskind. “The Science of using Science: Towards an Understanding of the Threats to Scaling Experiments,” *International Economic Review*, 61(4), 2020.
- ❖ Al-Ubaydli, Omar, Lai, Chien-Yu & John A. List. “A Simple Rational Expectations Model of the Voltage Effect,” NBER, 2023.
- Al-Ubaydli, Omar, Min Sok Lee, John A. List, Claire Mackevicious, and Dana Suskind. “How Can Experiments Play a Greater Role in Public Policy? 12 Proposals from an Economic Model of Scaling” *Behavioral Public Policy, Special Issue*, 5(1), 2021.
- List, J.A., D.L. Suskind, L. Supplee, edited book: “The Scale-Up Effect in Early Childhood and Public Policy,” Routledge, 2021.
- Supplee, LH, Ammerman, RT, Duggan, AK, Suskind, D, & List, J.A., “The Role of Open Science Practices in Scaling Evidence-Based Prevention Programs,” *Prevention Science*, 23(6), 2022.



THE VOLTAGE EFFECT

How to Make Good Ideas Great
and Great Ideas Scale

JOHN A. LIST

Internationally bestselling co-author of *The Why Axis*

Laws of Scaling

The 1st Law of Scaling:

VOLTAGE EFFECT: Benefit-cost profile changes (usually depreciates) when you move from the small to the large

Analogy: it is the scaling up (higher voltage) that enables ideas to move to many people and locations

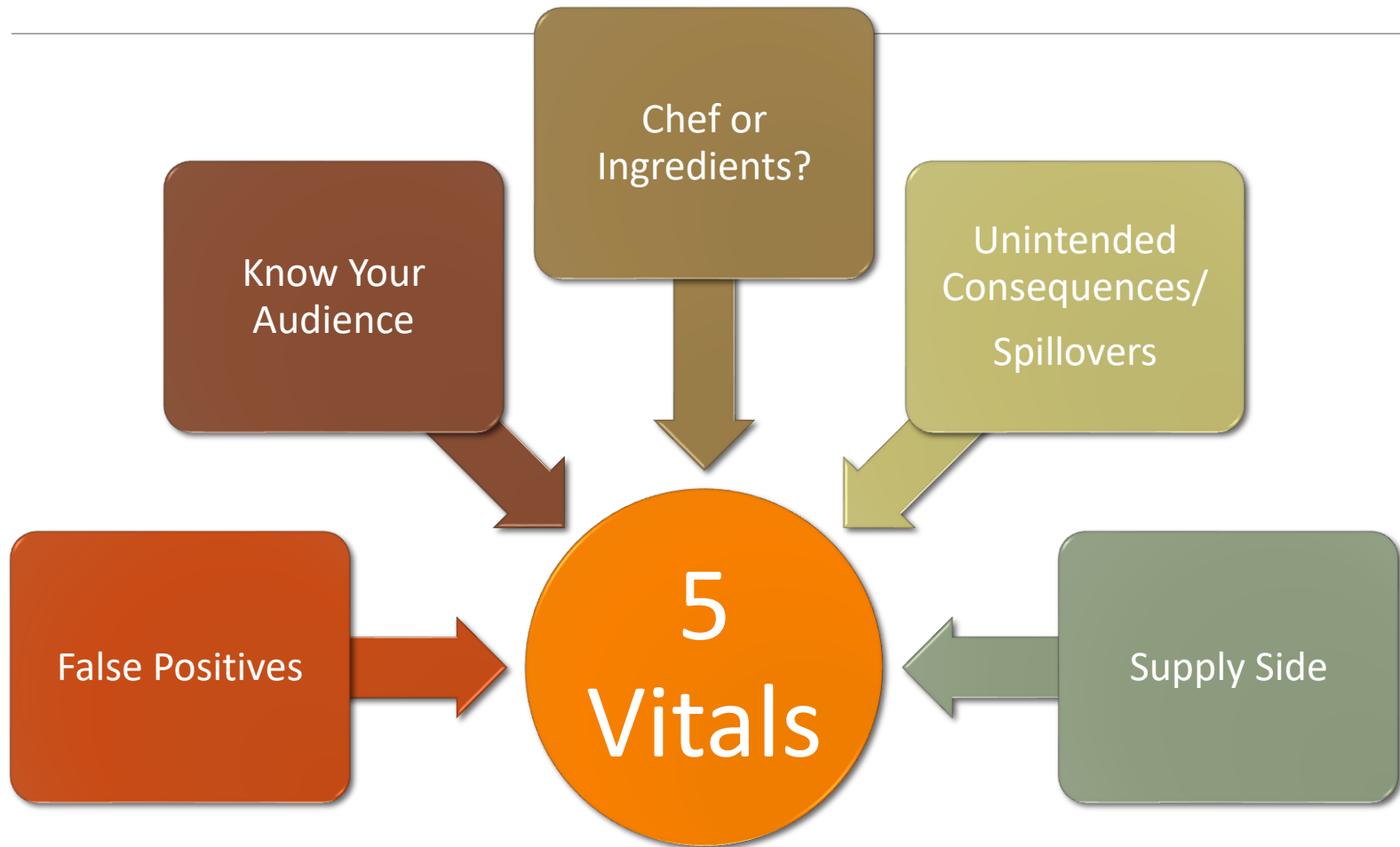
Laws of Scaling

The 2nd Law of Scaling:

This is not about a “silver bullet”: idea must possess the 5 Vital signs before one can have scaling confidence

These 5 Vitals provide insights into breadth and depth of your idea or policy

5 Vitals of Scaling



Vital #3

Representative Situation?

Horizontal Scaling meets Vertical Scaling

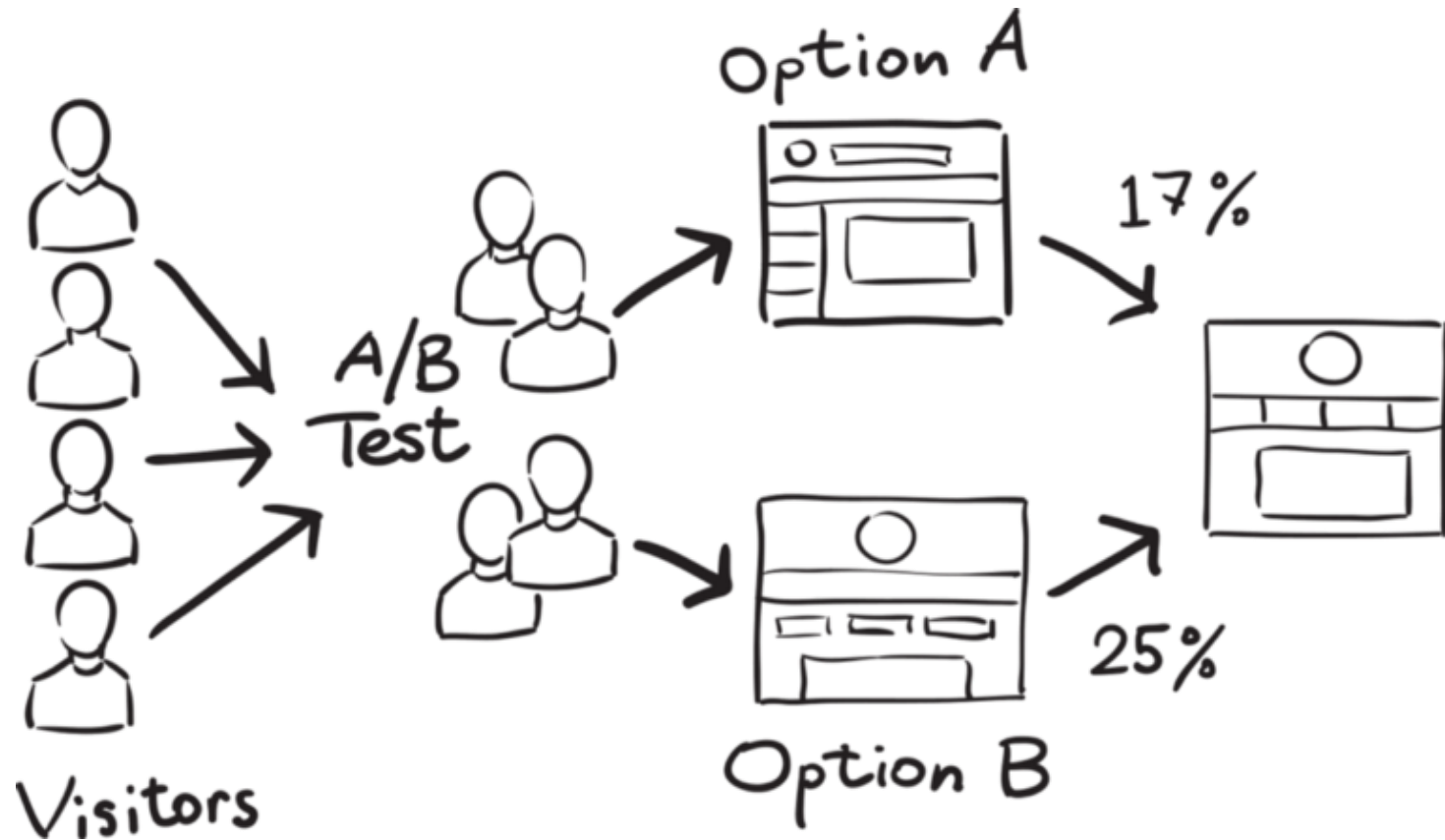
- Spread to other situations and settings
- Must scale across input/output markets (horizontal) and within an input/output market (vertical)

Humans don't scale well: hiring the 30 best teachers might be a good test of horizontal scaling but you learn little about vertically scaling (beyond 30 teachers)



We are asking the
wrong question

Today's Approach: do A/B with a dash of respondent heterogeneity





Efficacy tests are the wrong approach if you want to change the world at scale

How do we tackle what is necessary?



Option C Thinking = Policy-Based Evidence

Key Lesson: block on key situational features that can provide insights about critical features the idea will face at scale just like we block on individual characteristics today

→ Policy-Based Evidence

Option C Thinking

Envisioning how the scaled situational features depart from the experimental setting and designing appropriately

List, JA. “Optimally Generate Policy-Based Evidence Before Scaling,” *Nature*, 626, pp. 491-99, 2024.



Introducing Option C Thinking to A/B tests

- The Option C analogy is meant to take the initial discovery process from one of focusing purely on the details of an efficacy test to engaging in a bigger picture view, including the examination of:
 - What new constraints the idea will face at scale.
 - What key factors can impact scaling.
 - Whether the mediation paths and moderators are in place at scale.
 - Incredible spot for EXPERIENCE, theory, AI, and the like to enter
- At CHECC, we designed our field experiment to examine whether our curriculum could work with teachers who have varying abilities.
- How can we leverage AI?

ChatGPT Instructions

1. Train ChatGPT on what Option C Thinking is using “Optimally generate policy-based evidence before scaling” (List, 2024)
2. Feed ChatGPT the following prompt:
 1. The aim of this project is to generate 3 Option C Thinking ideas for the experiment
 2. I provide:
 - The research question
 - A description of the setting
 - The experimental population
 - The intervention(s)
 - The scaling aims
 - Any additional information, if necessary

ChatGPT Prompt for CHECC

Prompt to ChatGPT

1. What is the impact of preschool on future achievement?
2. Early childhood program in the south suburb of Chicago, Chicago Early Childhood Center
3. Sample of children from low-income families
4. Children are randomly assigned to receive instructional preschool or be in the control group
5. We would like to scale this program across the U.S.
6. No additional information

Option C Thinking Idea #1

Teacher Pool & Resources

In the best-case scenario (A/B test), one might hire especially motivated, highly credentialed teachers or give them extra resources. At scale, it's far likelier one will have to hire from a standard pool of teachers with potentially less robust training, varied levels of experience, etc.

Option C Thinking Idea: recruit teachers under the normal district hiring process (i.e., limited salary caps, standard certifications)

Resources: provide typical classroom resources (e.g., standard teacher–student ratios, curriculum materials consistent with state mandates)

Rationale for this Option C Thinking Idea

This lets you test whether the preschool model still boosts achievement when teaching staff and resources reflect ordinary U.S. public-school settings rather than a “dream team” that might have been assembled for an A/B test

Option C Thinking Idea #2

Large Classroom Size

At scale, classrooms might be larger than in a classic A/B test or one must replicate the intervention across many classrooms or multiple districts. Larger class sizes can reduce individualized attention and could impact results

Option C Thinking Idea: Instead of capping class sizes at, say, 15 children per teacher, have a larger ratio (like 20–25 per teacher), or operate multiple classrooms concurrently

- **Standard Real-World Facilities:** use typical classroom space, with minimal renovations or specialized learning spaces

Rationale for this Option C Thinking Idea

By replicating the more crowded conditions often faced nationwide, one will see how the effectiveness of the preschool program holds up when teacher – child ratios and facility constraints are closer to reality

Option C Thinking Idea #3

Variation in Parental Engagement & Supports

In an A/B test, parents of low-income children may have self-selected into the program or been especially motivated. At scale, not all parents will engage equally (work schedules, transportation, competing life demands). This matters because parent involvement can strongly influence outcomes

Option C Thinking Idea: compare outcomes if one does not provide any additional incentives or parent-focused resources

- **Diverse Scheduling & Communication Approach:** offer the usual day-to-day communication (e.g., standard teacher notes, a typical app or email updates) without extra home visits or specialized parent sessions

Rationale for this Option C Thinking Idea

This condition tests whether the preschool model still confers learning gains for children whose parents participate in only typical ways

A Few Further Thoughts: Representativeness of the Situation

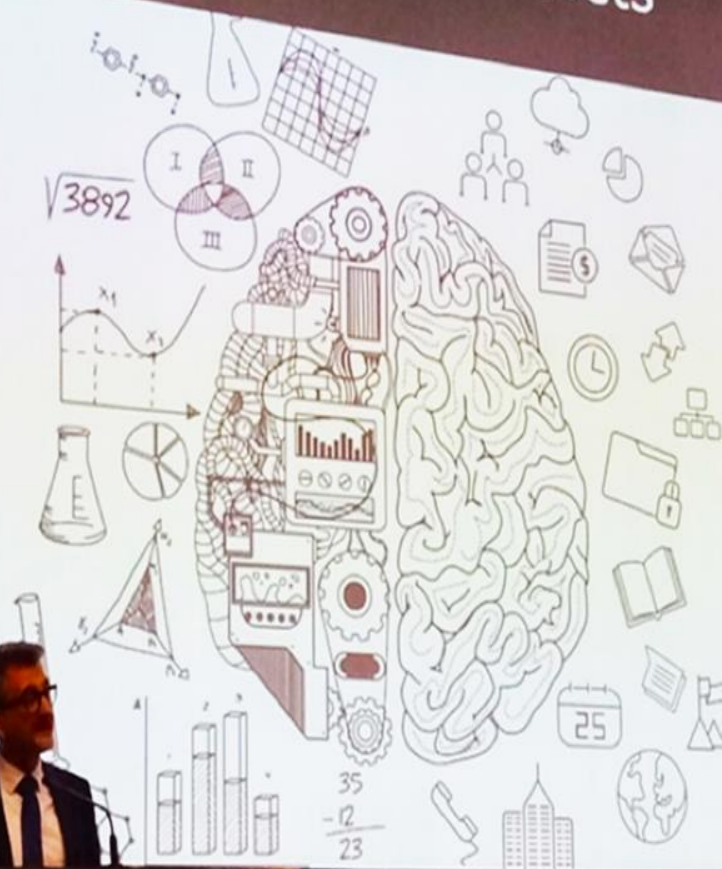
4 Insights

- ❑ Fidelity is increased if facilitators understand the “whys” behind the intervention effect (understand mediators and create the simplest path(s) to maximize impact)
- ❑ Technology should be used whenever possible (standardization, correct dosage, correct program, etc.)
- ❑ Policymaker must understand negotiables and non-negotiables when scaling (demand these from scientist)
- ❑ Using the original scientist on the implementation team can enhance fidelity

Different mindsets

Scientist

Policy-maker



European
Commission

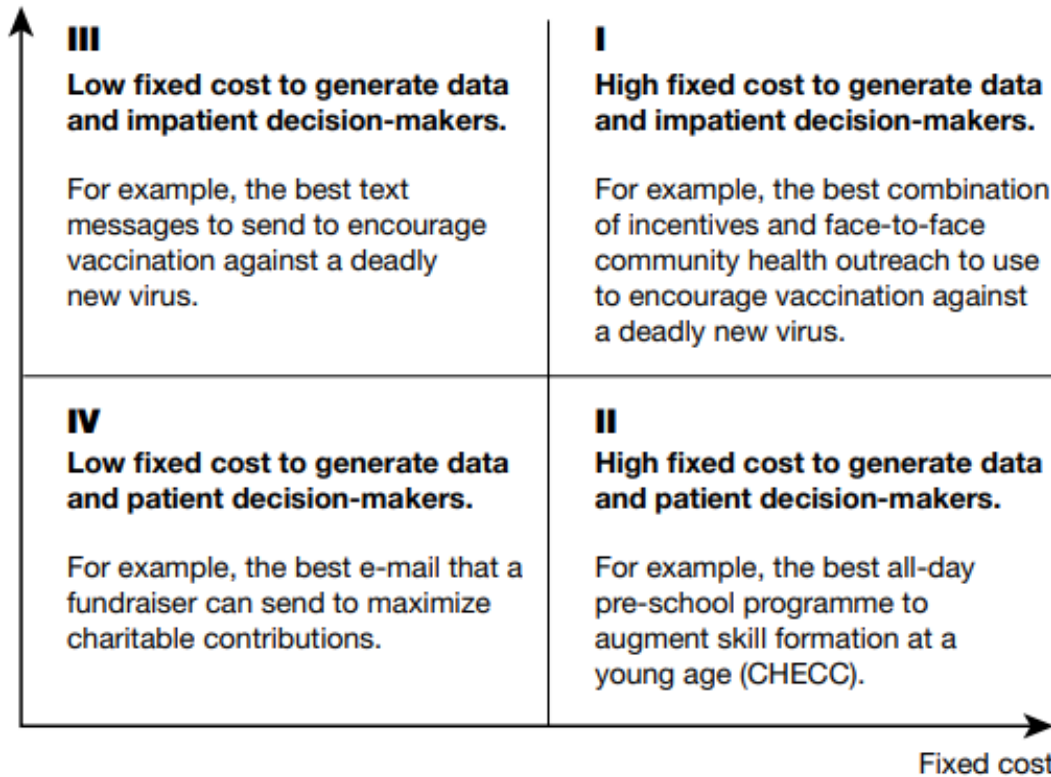
A New Landscape



- Yesterday's economist visited the pin factory to scribe general economic theories.
- Today's economist models natural disruptions that affect the pin factory and occasionally generates data from the pin factory to test economic theories.
- Tomorrow's economist (ALL OF US) will surgically control the assignment mechanism, in an ethically responsible way, to provide causal evidence of the inner workings of the pin factory *and* simultaneously generate data necessary to understand if learnings from that environment will transfer to other environments, be them spatially, temporally, or scale differentiated.

Introducing HFIDs

Decision-maker
impatience



What types of ideas are most important for the researcher to **backward induct**?

Ceteris paribus, the strongest case can be made for **ideas in region I**.

List, JA. "Optimally Generate Policy-Based Evidence Before Scaling," *Nature*, 626, pp. 491-99, 2024.