



UNIVERSITY
OF WARSAW



FACULTY OF
ECONOMIC SCIENCES

WORKING PAPERS

No. 39/2026 (533)

SIX BLIND MEN AND THE ECONOMICS-ELEPHANT: WHY COMPUTATIONAL POWER CANNOT CURE METHODOLOGICAL BLINDNESS

TOMASZ KOPCZEWSKI
IANA OKHRIMENKO
JAN LISICKI

WARSAW 2026

ISSN 2957-0506



Six Blind Men and the Economics-Elephant: Why Computational Power Cannot Cure Methodological Blindness

Tomasz Kopczewski^{1}, Iana Okhrimenko², Jan Lisicki¹*

¹ *University of Warsaw, Faculty of Economic Sciences*

² *Lazarski University in Warsaw, Department of Econometrics*

^{*} *Corresponding author: tkopczewski@wne.uw.edu.pl*

Abstract: Economics rewards precision, but precision within a model does not establish what a result means, whether the represented object is economically admissible, or whether the conclusion travels to the world. This methodological essay builds a classroom prank on the Indian parable of the blind men and the elephant. Students solve slightly different calibrations of a two-sector Leontief model, reduce the answers to policy signs and vote; every system is solvable, yet thirty of the hundred coefficient pairs fail the Hawkins–Simon productivity condition. Six blind men extend the lesson to calculation, point equilibrium, binary inference, identification, aggregation and computational dynamics. In each case the well-behaved statistic is not the one that answers the economic question, and its good behaviour is what makes it persuasive. Artificial intelligence lowers the cost of building alternative models towards zero; the scarce input is the disposition to ask what a successful result is not showing.

Keywords: economic methodology, model instability, Hawkins–Simon condition, statistical significance, classroom demonstration

JEL codes: A22, B41, C12, C63, C67, D51

1. Introduction: The Parable, the Prank, and the Thesis

“Do not take what I say as if I were merely playing....”
—Socrates, in Plato’s *Gorgias* (ca. 380 B.C.E./1925, 500b–c)

Teaching economic methodology is difficult in introductory and advanced courses alike. Students typically encounter methods as instruments for obtaining answers, while the conditions under which those answers deserve an economic interpretation often remain implicit. Even mathematics education has long distinguished instrumental command of rules from relational understanding of why they work (Skemp 1976). It has also shown how technically successful instruction can produce fragmented, procedure-centred knowledge and a suspension of realistic sense-making (Schoenfeld 1988; Verschaffel et al. 2000). Economics adds its own pressure towards formal closure, in which correct technique can displace substantive argument (McCloskey 1998). A conventional methodological warning can be absorbed as one more rule to remember. Sometimes an unconventional intervention is needed, one that leaves a durable trace and gives students permission to step outside their methodological cocoon.

The Indian parable commonly known as ‘the blind men and the elephant’ is a well-worn story. Six blind men encounter different parts of a great beast and mistake limited tactile evidence for the whole truth. The man at the tusk declares the elephant a spear. The man at the tail insists it is a rope. The man at the leg argues it is a tree. Their perceptions are locally accurate: a leg can indeed feel like a tree. The error begins when a locally true description becomes a theory of the whole elephant. Economic models may work in much the same way. One reveals an equilibrium, another identifies a causal effect, and another generates an emergent pattern. Each may be rigorous within its own field of vision.

The parable is often read as a story about too few observations, which would make it a story that more data could end. But the men are blind, not merely few. Their problem is not only coverage: the parts arrive without their relations, and touch cannot provide what sight would reveal. No number of additional tactile observations yields colour, an outline at a distance, or the placement of each part within a visible body. Two limits therefore survive any increase in computational power: the parts do not compose by themselves, and the instrument fixes what can count as evidence at all.

Chang (2014) uses the parable to describe competing schools of economic thought. This essay redirects the metaphor towards the mathematical tools of the modern economist.

Its blind men are competent technicians whose methods answer real questions and whose field of vision is necessarily narrow. In some cases, tunnel vision is the more precise term: the economist sees one tractable fragment with exceptional clarity while the surrounding object disappears from view. Calculation, equilibrium analysis, statistical inference, econometric identification, axiomatic aggregation, and computational dynamics each reveal something important. Methodological blindness begins when the output of one procedure is treated as a sufficient description of the economics-elephant.

Our teaching strategy begins with a prank. A classroom exercise, presented as a speed test in mathematical competence, allows students to experience the problem before it is named. They solve slightly different calibrations of a two-sector input–output model, report only policy signs, discover sharp disagreement, and vote on a common recommendation. Every assigned system yields an algebraic solution, yet the full grid crosses the boundary between a productive and a non-productive Leontief economy. The prank follows a Socratic sequence: a familiar procedure generates confidence, an apparently clear collective verdict breaks that confidence, and the shared confusion becomes material for methodological inquiry.

The debriefing introduces six blind men. Each represents a different point at which a useful analytical practice can become a stopping rule:

The Answer Getter treats calculation as completion: once a number appears, interpretation can stop. The Point Equilibrium Believer lets one conditional coordinate stand for the landscape on which it depends. The P-Value Voter converts graded evidence into a binary verdict. The Econometrician with Perfect Data trusts precision before identification, specification, and causal interpretation have been secured. The Axiomatist constructs a tractable whole by imposing enough structure for individual properties to survive aggregation. The Chaos Hunter restores time, feedback, nonlinearity, heterogeneous interaction, and emergence, then may mistake an internally generated pattern for an explanation of the external world. Each form of blindness grows from a real strength of the underlying method.

The common problem is therefore methodological rather than technical. A recurrent pattern can be stated compactly: the well-behaved statistic is not the one that answers the economic question, and its good behaviour is what makes it persuasive. More computing power can make each local result faster, smoother, and more precise while leaving its relation to the other five perspectives unresolved. The same applies, with greater force, to systems that can perform all six operations at once and narrate the result. The parable changes only when

the blind men compare their evidence and revise their accounts. McCloskey's (1998) account of economics as human argument captures the role assigned here to conversation: partial results are confronted with assumptions, alternative representations, and the world they are meant to describe. Section 9 asks what artificial intelligence changes about the cost of that comparison and what it leaves untouched.

The classroom episode directly demonstrates parameter sensitivity, economic admissibility, information loss, and authority within a reporting procedure. The later sections use the experience as a bridge to broader methodological problems. The prank creates a concrete reason to ask about multiple equilibria, causal identification, aggregate instability, and deterministic chaos. Section 2 describes the classroom prank and its analytical structure; Sections 3-8 develop those questions with the literature and formal arguments appropriate to each case; Section 9 brings the partial perspectives into conversation, asks what artificial intelligence changes about its cost, and returns the argument to the classroom in which it began.

The arguments can also be explored rather than only read. A live beta version of the companion dashboard allows readers to reconstruct any of the hundred classroom assignments and then vary the parameter neighbourhood, significance threshold, regression geometry, degree of heterogeneity, and law of motion used by the six blind men. It is available at https://microeconomics.shinyapps.io/six_blind_men/. The application is deliberately maintained as a living laboratory: the address points to its latest version, and each of its pages states whether it reproduces the classroom design or adds a synthetic extension. Readers are invited to move the controls, search for counterexamples, and ask where each conclusion ceases to hold.

2. The Classroom Prank: When the Model Becomes Background

The prank is not a fixed part of the curriculum. It has been used in several forms, mainly in advanced or applied microeconomics and in master's and doctoral seminars. It has been run more than twenty times; answers were not recorded systematically, so the classroom observations reported in this essay are the authors' observations from teaching, not data. Its wording requires no prior knowledge of economics. The Leontief model supplies the economic setting, but the task confronting students is simpler: they are asked to solve a system of two linear equations. That distinction is central to the prank.

The class begins like a conventional lecture, preferably during the first meeting of the course. The instructor adopts the role of a strict guardian of technical competence

and announces a short calculation test. Students may use calculators, but they must record the time on their smartphones. The announced purpose is to assess how much mathematical revision the group will require.

The instructor begins:

“You know that we will use mathematical tools throughout this course. Before deciding how much time we should devote to calculation, I would like to assess both your competence and your speed.

Consider a stylised economy with two sectors. Sector 1 produces goods, by which we mean everything outside the electricity sector. Sector 2 produces electricity. The sectors depend on one another. Goods require electricity, while electricity production requires equipment, maintenance, fuel, and electricity itself.

The economy is facing an energy crisis. Final demand for goods is expected to rise by 40 billion. An efficiency programme is expected to reduce final demand for electricity by 10 billion. The ministry wants to know by how much the gross output of each sector must change.”

The instructor then shows how the situation can be represented in a model:

A convenient representation of this situation is the Leontief input–output identity $x = Ax + d$ (Leontief 1986). The vector x records gross outputs, d records final demand, and A is the technology matrix. Its entry a_{ij} gives the amount of product i directly required to produce one unit of product j ; Ax records the intermediate demand generated by production itself.

For the two-sector economy used in the exercise, the technology matrix is:

$$A(e_1, e_2) = \begin{pmatrix} 0.60 & 0.54 \\ 0.10 + e_1 & 0.85 + e_2 \end{pmatrix}$$

Two coefficients are treated as uncertain. For the purposes of the exercise, each student receives a private pair of values determined by the final digits of the student ID:

$$e_1 = \frac{\text{last digit} - 5}{100}, \quad e_2 = \frac{\text{second-to-last digit} - 5}{100}.$$

To discourage copying, the instructor assigns a different pair to each student. Each parameter therefore ranges from -0.05 to 0.04 . With the technology matrix fixed, the required change in gross output must accommodate the change in final demand:

$$(I - A)\Delta x = \Delta d, \quad \Delta d = \begin{pmatrix} 40 \\ -10 \end{pmatrix}.$$

In other words, the required output changes can be obtained by solving the following system:

$$\begin{aligned} 0.40\Delta x_1 - 0.54\Delta x_2 &= 40 \\ -(0.10 + e_1)\Delta x_1 + (0.15 - e_2)\Delta x_2 &= -10. \end{aligned}$$

Operationally, this is now the whole assignment. Students are not asked to choose the model, defend its assumptions, inspect the coefficients, or decide whether the resulting economy is possible. They are asked to solve two equations quickly and report the signs of the two unknowns.

This reduction is part of the trap, but it also reproduces a familiar form of economics teaching. The model introduces the story and legitimises the exercise; the calculation becomes the visible work and the assessable achievement. A student may perform the mathematics correctly without ever having to ask what economic object the equations describe. The Leontief model has become background, while solving the system has become a way of teaching economics.

The instructor then adds the crucial misdirection:

“Your coefficients differ slightly from those received by your neighbours, so your numerical answers will not be identical. The differences are small, however, and should not affect a consistent interpretation across the group.

Interpretation is simple. A positive result means that the sector must expand; a negative result means that it must contract; an exact zero should be recorded as 0. When you finish, come to the blackboard and write only the signs of your recommendation, for example (+, −) or (+, 0), and the time required for your calculation.

Do not communicate with one another or ask questions during the test. I am assessing both individual performance and the group as a whole. You have four minutes.”

Within two or three minutes, the blackboard begins to fill. Some students report that goods production should rise while electricity production should fall. Others find that both

sectors must contract. A third group obtains expansion in both sectors. Students whose last ID digit is five obtain an exact zero for the change in electricity output.

The disagreement initially resembles a collection of arithmetic errors. Students recheck their calculations. Those waiting to write can already see others' submitted answers. A result that differs from the first entries on the board may produce hesitation, another calculation, or quiet suspicion that something has gone wrong.

Sequential reporting changes the informational environment, but the exercise does not identify the social mechanism. A student may be responding to authority, embarrassment, rational updating, elementary error checking, or simple confusion. Labelling the episode as conformity or an information cascade would claim more than the design can establish.

When the four minutes have elapsed, the instructor examines the board and escalates the exercise:

“You can see what time pressure has done. This is only a two-sector model with four coefficients, yet the group cannot agree on the required policy because of calculation errors.

Let us therefore approach the problem statistically. Some of you probably made mistakes and some did not. We will see whether the group can identify the correct recommendation collectively.”

For the vote, the instructor performs an additional act of compression. The recommendations $(+, -)$ and $(+, 0)$ are combined into one policy category: goods production should expand while electricity production need not expand. The resulting alternatives are $(+, \leq 0)$, $(-, -)$, and $(+, +)$. Students are invited to reconsider their positions after seeing the board and then vote by raising their hands. Under this plurality rule, the recommendation with the most votes wins; a tie is recorded as no unique winner.

The procedure has now become openly absurd. Students have solved different parameterisations of the economy, yet the class is invited to manufacture a single recommendation. The voting rule discards magnitudes, the distinction between contraction and no change, the parameters behind each result, and the question of whether every assigned economy is admissible.

The vote has a deceptive stability. Across the complete grid of one hundred assignments, 57 produce the compressed recommendation $(+, \leq 0)$, 26 produce $(-, -)$, and 17 produce $(+, +)$.

In ten-person classes in which each student receives a different pair, drawn at random from this grid, $(+, \leq 0)$ is the unique plurality winner with probability 81.40 per cent: it receives more votes than either rival, and a tie does not count as a win. Its probability of securing a strict majority, more than half of all votes, is 55.82 per cent. The aggregation rule tends to deliver a clear answer. Its clarity is precisely the problem.

2.1. The Reveal

The instructor stops the test and addresses the class:

“First, I owe you an apology. I have just pranked you.

I allowed a sequence of apparently reasonable actions to carry us towards an absurd conclusion. Such a device would be cruel if its purpose were to make the teacher look clever and the students look foolish. I would not put you through this for the pleasure of watching you struggle.

If any of you feels that I made a fool of you, that is a fair reaction. I used your trust in me as the person conducting this class. The prank has educational value only if I reveal the mechanism and accept responsibility for creating it.

Look again at the board. The disagreement was not produced mainly by a lack of mathematical competence. Most calculations are correct for the coefficients assigned to the person who performed them. I built the disagreement into the task and then assured you that the differences would not matter.

I announced a test of competence. I gave you four minutes. I prohibited questions and communication. I said that I would assess you. I presented the task as simple and routine. These instructions created a division of labour. I supposedly knew why the calculation made sense. Your role was to perform it quickly and correctly.

Calculation was rewarded. Questioning the object of calculation was costly. Refusing to solve the system could have looked like an admission that you did not know how. Interrupting the test could have appeared disruptive. Trusting the professor seemed reasonable.

I said “calculate”, and you calculated. I said “vote”, and you voted. My authority made a sequence of unreasonable actions appear reasonable.

We ended by asking the class to produce one policy from answers to different models.

Now let us inspect the models themselves. Notice where the prank began. It did not begin with the vote. It began when the economic model was allowed to retreat into the background and the system of equations became the task. You could therefore be technically successful without asking whether the coefficients described a possible economy. The story supplied the setting; calculation supplied what could be timed, compared, and graded.

Some coefficient pairs assigned in this exercise do not describe a productive input–output economy. Across the complete grid, this happens in thirty cases out of one hundred. I can identify the corresponding cases from the digits that generated your parameters.

More precisely, these matrices violate the Hawkins–Simon productivity condition (Hawkins and Simon 1949). Imagine asking the economy to deliver one additional unit of final output. Producing it requires intermediate inputs from both sectors; producing those inputs requires another round of inputs, and that round requires another. In a productive economy, the successive rounds become smaller and the chain converges. In a non-productive specification, the sectors use too much of one another's output for the chain to settle: production is absorbed by the attempt to reproduce the inputs needed for production itself. The equations may still return a pair of numbers because the matrix is invertible. The appearance of a solution therefore does not certify that the technology can support final demand.

Now look at the winning recommendation. In the complete grid, fifteen of the 57 results classified as $(+, \leq 0)$ come from coefficient matrices that fail the productivity condition. They look exactly like the other votes after we remove the numbers and retain only a policy category. The same reporting rule also absorbs the ten exact-zero results into the same category as contraction. Once values become signs, both economic impossibility and an economically meaningful distinction disappear from view.

There was another warning available to you. For one assignment, an increase of 40 billion in final demand for goods is associated with a required change

of 5,500 billion in gross goods output. A result of that magnitude should interrupt the procedure. It should provoke questions about the coefficients, the multiplier, and the distance of the system from singularity. Instead, our reporting rule transformed it into an ordinary plus sign.

This is the first purpose of the prank. Science cannot consist in obedience to authority. A calculation does not become meaningful because a professor has ordered it. Mathematical notation does not guarantee an economically coherent model. A plurality vote cannot recover information that the reporting procedure has discarded.

Scientific education should give you permission to interrupt a procedure when its object stops making sense. You should be able to ask: What exactly are we calculating? What do these coefficients mean? Which values are economically admissible? How large is the implied multiplier? Why should these parameterisations be combined? What has been lost by reducing the answers to signs?

I deliberately made these questions difficult to ask. The resulting failure was systemic. It was built into the incentives, the time limit, the reporting rule, and the distribution of authority. Competent people can perform every assigned operation correctly and still participate in a procedure that produces little understanding.

Distrusting the professor would be a small lesson. The harder task is to recognise the same mechanism when no professor is present.

Researchers calculate because a report requires a number. We estimate because a paper requires a coefficient. We test because a publication requires a result. We rank because an institution requires an evaluation. We forecast because someone requires a date and a point value. Every operation may be technically competent. Routine can gradually become its own justification.

Then we stop asking the fundamental questions. What are we doing? Why are we doing it? Which assumptions carry the conclusion? What information did we discard to obtain a clear answer? Does the result describe an economic object, or merely the procedure that generated it? What observation would make us reconsider our interpretation?

Critical thought must therefore be directed towards our own work. It is easy to criticise another researcher's assumptions or another school's method. It is much harder to notice the moment when our own competence becomes routine and routine becomes blindness.

The hardest authority to question is often the one that speaks in our own voice and says: I know this method, therefore I know what I am doing.

We will return to this blackboard throughout the course. When we discuss equilibrium, statistical inference, econometric identification, aggregation, dynamic instability, and computational models, we will return to what happened here. The board will not answer all those questions. It gives us a shared experience from which to begin asking them.

I do not want this prank to teach you to fear mathematics. Mathematics becomes more valuable when a successful calculation opens the next question instead of closing the inquiry. The purpose of this course is to learn how to ask that next question."

The exercise is not presented as human-subjects research, yet norms governing research deception provide a useful minimum ethical benchmark for its use. Deception must be justified by prospective educational value, must not expose students to severe distress or other reasonably foreseeable harm, and must be followed promptly by disclosure, debriefing, and a genuine opportunity to object (American Psychological Association 2017, Standards 8.07–8.08; Ortmann and Hertwig 2002). The apology is not theatrical closure; it is part of the method. The instructor should also make clear that performance in the prank is not graded and that students may decline further use or discussion of their responses.

2.2. The Anatomy of the Trap

The calculation represents a real planning problem. Input-output models are used to trace how an energy shock, an investment programme, or a change in household demand propagates through supply chains. Gross output must cover final demand and the intermediate inputs used in producing that output. The two-sector economy is stylised, but its question is literal: after the goods sector has supplied the electricity sector and the electricity sector has supplied the goods sector, is there enough production left for final users, and how much gross output is required to deliver it?

The determinant of the system matrix is:

$$D = 0.40(0.15 - e_2) - 0.54(0.10 + e_1).$$

The solutions can be written as:

$$\Delta x_1 = \frac{40(0.15 - e_2) - 5.4}{D}, \quad \Delta x_2 = \frac{40e_1}{D}.$$

The second expression explains the exact zeros on the board. Every student whose last ID digit is five receives $e_1 = 0$ and therefore obtains $\Delta x_2 = 0$, regardless of the second parameter.

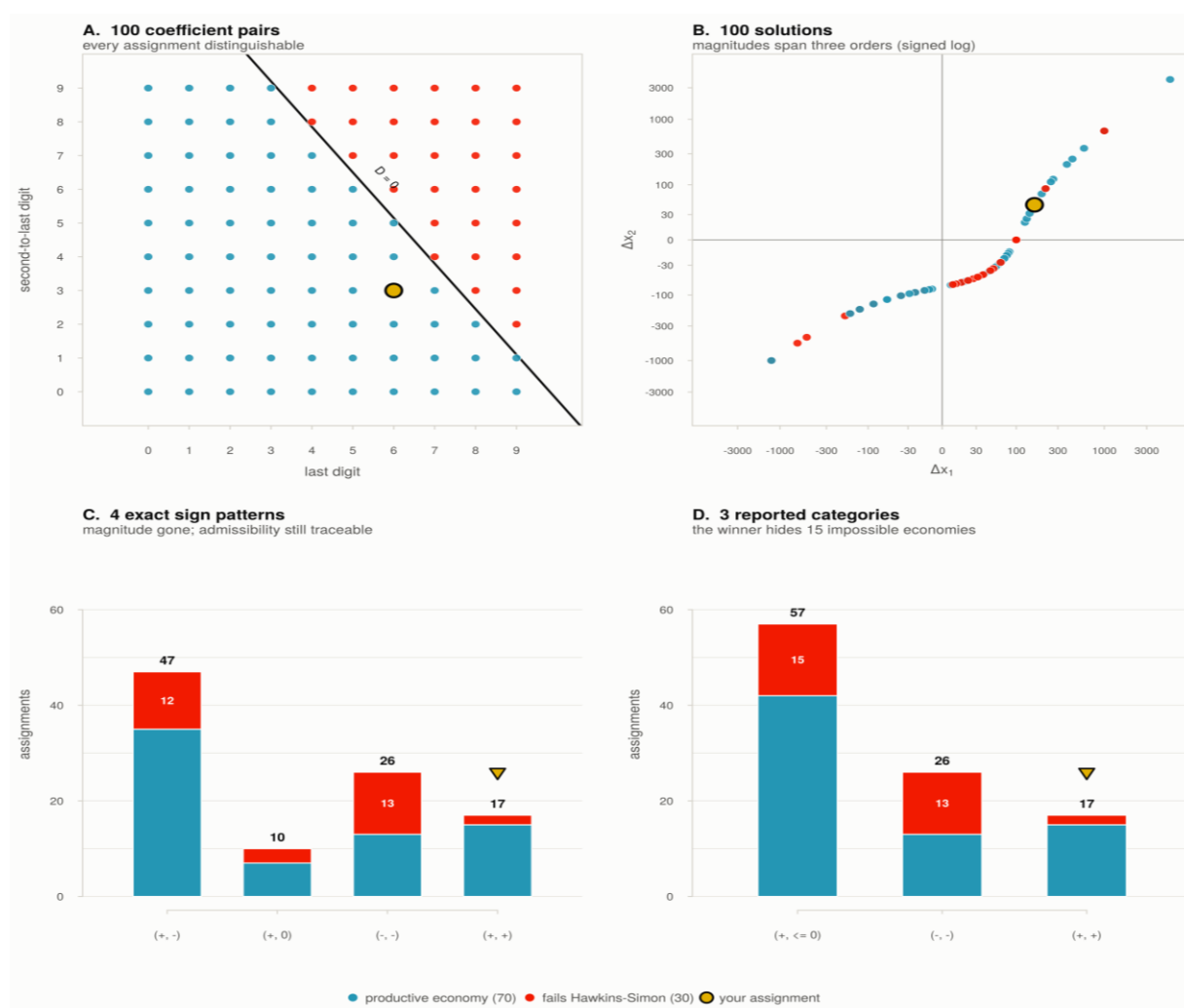
The Hawkins–Simon condition gives the denominator an economic meaning. With each additional round of production, the two sectors create further intermediate requirements. If these rounds shrink, the process converges and a finite non-negative gross-output vector can support any non-negative final-demand vector. If they do not shrink, the economy is technologically incapable of reproducing its inputs and still delivering final output. For this two-sector matrix, both one-by-one principal minors are positive throughout the grid, so the condition reduces to $D > 0$. The line $D = 0$ is not merely where division by zero occurs; it is the boundary between a productive Leontief economy and a coefficient system that cannot sustain final demand.

None of the hundred assignments makes D exactly zero, so all hundred systems are algebraically invertible. The grid nevertheless passes close to the boundary. The 2-norm condition number of $I - A$ has a median of approximately 30.5, reaches 1,268, and exceeds 100 in seventeen assignments. Across the same grid, Δx_1 ranges from $-1,250$ to $5,500$, while its median absolute value is approximately 58.8. The ratio between the largest and smallest non-zero absolute values is nearly 800. The largest absolute entry of the Leontief inverse has a median of approximately 34.6 and reaches 1,350. These are not large differences produced by large shocks; they are enormous differences generated by small changes in technical coefficients.

Thirty parameter pairs lie on the non-productive side of the boundary. Their equations remain solvable because $D \neq 0$, yet they do not represent productive Leontief economies. This is the first paradox: algebraic solvability survives after economic admissibility has failed. The second appears near $D = 0$: the calculation returns an exact-looking number precisely where the result is most sensitive to coefficient uncertainty. A requirement of 5,500 billion is produced by the same procedure that produces an ordinary value, although it should be read as a warning about the model rather than as a plausible production plan.

Figure 1 follows the same one hundred coefficient assignments through four descriptions. Panel A preserves the coefficients and shows the side of the Hawkins–Simon boundary on which each economy lies. Panel B preserves the algebraic solutions and their magnitudes. Panel C discards magnitude and retains only four sign patterns. Panel D compresses exact zero together with contraction and produces the three categories used in the vote. The last two panels reveal two further paradoxes. An exact zero disappears when no change is grouped with contraction, and the apparently stable winner $(+, \leq 0)$ contains fifteen economies that fail the very condition required for economic interpretation. The answer becomes clearer as the economic object becomes less visible. The highlighted assignment, with last digit 6 and second-to-last digit 3 ($e_1 = 0.01$, $e_2 = -0.02$), is reproduced on the dashboard page ‘The prank’; Panel B uses signed-log axes only because the values span three orders of magnitude.

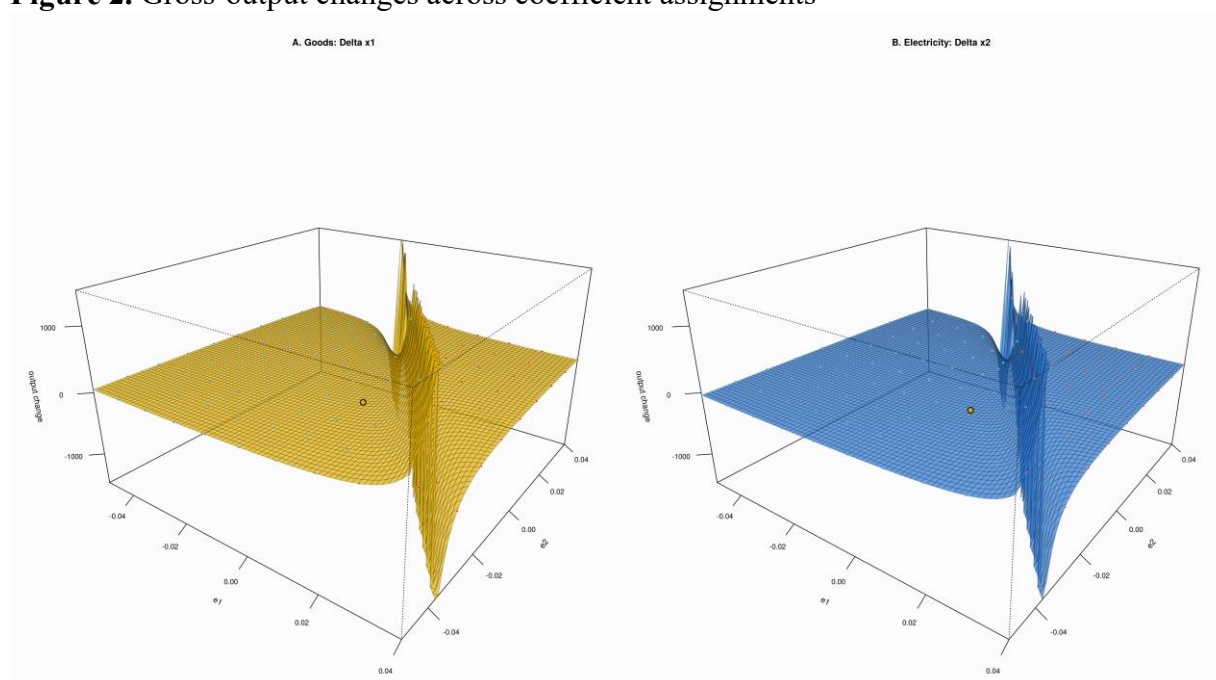
Figure 1. Four descriptions of the same hundred classroom economies



Note: Teal: productive matrices; red: matrices failing the Hawkins–Simon condition; gold: the highlighted assignment ($e_1 = 0.01$, $e_2 = -0.02$). Panel B uses signed-log axes.

Figure 1 shows where information is lost; Figure 2 shows why neighbouring assignments can produce radically different values. The continuous solution surfaces for Δx_1 and Δx_2 have a steep fold near $D = 0$. The discrete grid allows each student to locate the assigned economy, while the surfaces reveal the surrounding landscape. Together they distinguish three questions: Does an algebraic solution exist? Is the coefficient matrix economically productive? How stable is the numerical result to plausible changes in the coefficients? The surfaces are clipped at $\pm 1,500$ for display, so the gap in each panel marks the neighbourhood of $D = 0$.

Figure 2. Gross-output changes across coefficient assignments



Note: Panel A: Δx_1 (goods); Panel B: Δx_2 (electricity). Colours as in Figure 1. Surfaces clipped at $\pm 1,500$.

The disorder on the board arose from a static linear system. Near-singularity is not deterministic chaos. Chaos requires a law of motion and a specific form of sensitivity through time. The distinction matters because Section 8 later asks what changes when feedback and dynamics are introduced.

The two-sector calibration is deliberately stylised, but the paradoxes are not classroom inventions. Applied input–output analysis uses estimated technical coefficients to calculate production requirements after energy, demand, trade, or supply-chain shocks. The same logical mistakes can arise when an invertible matrix is mistaken for a feasible economy, an enormous multiplier is read as a forecast rather than as a diagnostic warning, and a stable sign is treated as a robust policy conclusion. The prank magnifies these risks so that students can inspect them before meeting them in larger empirical models. Its intuitions become the starting point

for the methodological warnings about equilibrium, inference, identification, aggregation, and dynamics developed in the sections that follow.

2.3. From the Blackboard to the Six Blind Men

After the analytical debriefing, the instructor tells the class the parable of the six blind men and the elephant. Each man touches a different part of the animal and describes it correctly: a leg is like a tree, a tusk like a spear, and a tail like a rope.

The instructor presents the prank and the parable as the point of departure for methodological discussion. Their connection lies in the structure of the students' experience. Each student solves a different parametrisation and leaves the calculation with a different picture of what 'the economy' requires. Each picture follows from locally valid reasoning, but none contains the whole parameter landscape. Placing the answers on one blackboard does not reconstruct it. The reporting rule and the vote compress these partial views further, turning different calibrations, magnitudes, and degrees of economic admissibility first into signs and then into a collective verdict.

In the sections that follow, we call the recurring stopping rules mathematical blindnesses in economics. They arise when technical success allows a correct local result to stand for the economic object as a whole, so that we stop asking what the model cannot see and what its assumptions exclude or make invisible.

The six figures introduced in Section 1 mark six places at which inquiry can close too early: the P-Value Voter repeats the compression performed by the classroom vote, and the Chaos Hunter may mistake static sensitivity for chaos.

Each specialist touches a real part of the problem, and mathematics makes those parts visible with precision. But precision can become permission to stop. The instructor asks the students to read this essay and announces that the prank and the parable will return throughout the course whenever they work with mathematical, statistical, and computational models in economics. Their purpose is to leave a durable trace in the students' thinking. The blackboard begins the conversation; it cannot complete it. Whenever a technically successful result appears, that trace should prompt the next questions: What does this model allow us to see? What lies outside its field of vision? What have its assumptions excluded or made invisible?

3. The First Blind Man: The Answer Getter

The prank first exposes a familiar ambition: get the answer. Each student receives coefficients that look like small variations around a common calibration. The system is solvable in every assigned case, and successful algebra supplies the expected pair of output changes. The warning appears elsewhere. Some magnitudes are extraordinary, and thirty of the hundred possible coefficient pairs fail the productivity condition of the input–output model. The decisive question is no longer whether an absurdity is visible in the answer. It is how the solver knows that an ordinary-looking answer belongs to an admissible economy.

The first blind man is an Answer Getter. He may calculate quickly, carefully, and correctly. His blindness begins when numerical output becomes a substitute for inference. The number supplies closure before anyone asks what it means, whether its magnitude is plausible, whether its sign is admissible, or whether the model has produced an object that can exist in the economic situation being represented. Skemp's (1976) distinction between instrumental and relational understanding captures part of this difference: knowing a rule that produces an answer is a different achievement from understanding why the rule applies and how the answer is connected to the problem.

The exercises used in this line of teaching deliberately make this stopping rule visible. In one task, a mechanically correct calculation produced a negative price in a setting where the economic mechanism and the stated assumptions gave that result no sensible interpretation. Negative prices can, of course, be meaningful in some markets. That exception makes the lesson sharper: admissibility comes from the represented mechanism and institutional setting, not from a universal rule that prices must be positive. In another exercise, the calculation produced a negative probability. The impossible value was nevertheless copied into the answer. The calculation had finished, so the reasoning had finished with it.

A particularly revealing objection came from a mathematically trained student who defended an analytically correct but economically inadmissible solution by arguing that the domain of the function had not been specified. The objection was mathematically impeccable. It exposed the boundary between solving an equation and modelling an economy. Formal mathematics can legitimately ask the modeller to specify the domain. Economic interpretation also supplies part of that domain because words such as probability, quantity, population, share, or price refer to objects with substantive restrictions that depend on the model and the institution being described.

Mathematics-education research documents a closely related suspension of sense-making. Schoenfeld (1988) showed that students can perform well in a carefully taught mathematics course while acquiring a fragmented view of the subject centred on procedures and correct answer forms. Verschaffel et al. (1994) found that pupils often ignored realistic constraints when ordinary arithmetic operations yielded an answer. Greer (1997) argued that this behaviour is encouraged by the culture of classroom word problems, where students learn to select an operation, manipulate the numbers supplied, and report the result. Reusser and Stebler (1997) described the corresponding social rule with a provocative title: every word problem has a solution.

The phenomenon is not confined to young pupils. Verschaffel et al. (1997) found a strong tendency among pre-service teachers to exclude real-world knowledge both from their own solutions and from their evaluations of pupils' answers. Verschaffel et al. (2000) synthesised this literature as a problem of suspended sense-making. Palm (2008) found that more authentic task conditions increased the use of real-world considerations, while superficial solution strategies and beliefs about the genre of school mathematics remained important sources of unrealistic answers. Kirkland and McNeil (2021) likewise showed that changing the design of a word problem so that students had to reason explicitly about uncertainty improved subsequent sense-making. The common lesson is methodological: the form of the task can train people to treat calculation as the expected endpoint.

Across repeated teaching applications, the same behavioural endpoint has appeared at very different levels of mathematical preparation, although the routes to it seem different. At one extreme, repeated time-pressured drill rewards rapid completion: identify the formula, insert the numbers, obtain the answer, move on. At the other, highly formal training can make the mathematical specification itself the entire contract of the exercise. Procedural fluency and sense-making are separate competencies, and success in the first does not guarantee the second.

A useful diagnostic is to remove a number rather than add a harder calculation. An exercise that says 'assume a plausible annual interest rate', 'choose a reasonable inflation rate', or 'use a realistic depreciation rate' can paralyse a student who is perfectly comfortable once the parameter is supplied to four decimal places. The missing operation is quantitative judgement. The solver must know an order of magnitude, a unit, a time horizon, an institutional context, and enough about the world to defend the assumed value. There may be several defensible numbers. The task asks for reasoning before calculation.

Economics has its professional version of answer-getting. Mirowski (1991) warned that concentration on calculation and data processing can narrow the field of vision, while McCloskey (1998) criticised a modernist routine in which formal procedure can displace substantive argument. Manski (2011) gives the problem a particularly useful name: incredible certitude. Policy analyses can report precise point predictions even when their conclusions rest on unsupported assumptions, fragile extrapolations, or leaps of logic. Numerical precision then exceeds the information that supports it.

Leamer's (1983) response was to make sensitivity itself part of the reported result. Instead of defending one preferred specification, the researcher should map assumptions into inferences and show where small changes reverse the conclusion. For Leamer, that mapping—not a single preferred estimate—is the central message of an empirical analysis. This is what the prank's parameter plane and solution surfaces are meant to teach. A single computed pair $(\Delta x_1, \Delta x_2)$ answers a conditional question. The map shows how heavily that answer depends on the coefficients and whether the point lies in the productive region of the model.

The temptation does not disappear among practising economists. After a crisis, recession, or forecast failure, a model can often be recalibrated until a nearby parameter value reproduces the event. That exercise becomes informative when the revised parameter was plausible before the event, can be independently identified, or corresponds to a mechanism supported by evidence. Otherwise the new value mainly shows that the model can be rescued after the fact. Caballero (2010) described a related pretense-of-knowledge syndrome in macroeconomics and criticised the confusion of precision inside a model with precision about the real world. His discussion of retrospective 'Cassandras' is especially relevant: seeing a crisis in one recalibrated trajectory says much less than establishing beforehand which mechanism and parameter region made that trajectory empirically credible.

The supplementary simulation is designed to break the Answer Getter's stopping rule. Students vary e_1 and e_2 , watch the solution move, and cross the line $D = 0$. A result now invites a sequence of questions: What are its units? Is its magnitude plausible? Does the coefficient matrix satisfy the productivity condition? How close is the system to a boundary at which the answer changes sharply? Which assumptions carry the result? What observation would make us reject the interpretation?

For the Answer Getter, the number closes the problem. Economic reasoning begins when the number is returned to the world from which the model abstracted it. That move prepares

the second blind man. Once a numerical answer has survived interpretation, the next question concerns its status: should one coordinate be allowed to stand for the landscape on which it depends?

4. The Second Blind Man: The Point Equilibrium Believer

When students reached the blackboard with contradictory results, they searched for the correct answer that others had missed through careless arithmetic. The reaction is understandable. Much of economic training is organised around well-specified problems that terminate in a solution, often called the equilibrium. For every fixed coefficient pair in the prank, that instruction is mathematically legitimate: the linear system has one algebraic point solution. The methodological warning begins when the point is read without the conditions that give it an economic interpretation.

Economics has good reasons to organise reasoning around such points. Hicks (1939) and Samuelson (1947) made comparative statics extraordinarily productive by comparing equilibrium positions before and after a change in an exogenous parameter. The path between them can be set aside while the direction and magnitude of the effect become transparent. Hahn (1973) described equilibrium as a central organising idea of economic thought, and it earned that description: it supplies a precise object around which otherwise interdependent plans, prices, choices, and constraints can be arranged.

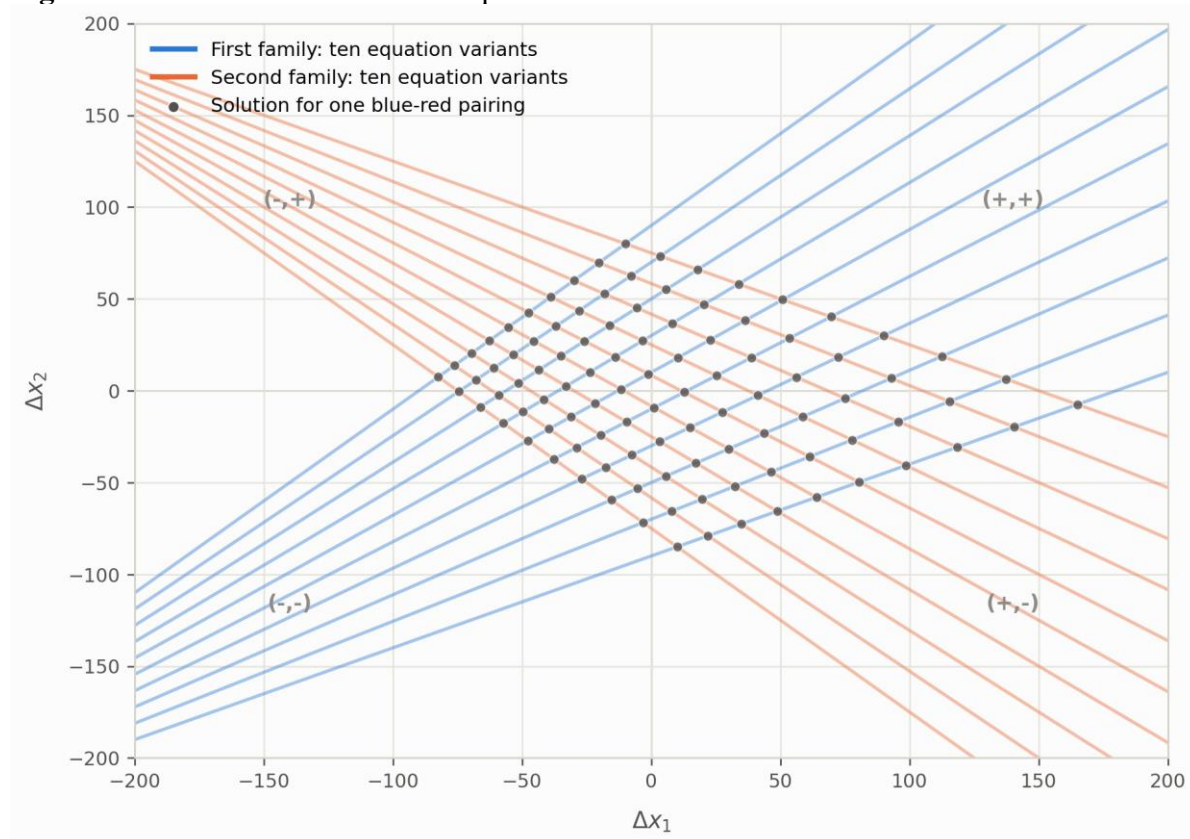
The problem is not the concept but the stopping rule it can create. Precision makes the object easy to report and its conditions easy to forget. The phrase ‘the equilibrium’ sounds self-contained even when the reported coordinate depends on a calibration, an admissible domain, a dynamic adjustment process, or a selection rule. The Point Equilibrium Believer mistakes the successful construction of one conditional object for the discovery of an unconditional fact.

4.1. A Point Is a Conditional Statement

Write the solution as $z(\theta)$: given a model, a parameter vector θ , an admissible domain, and any selection rule the model requires, this is the answer. Point thinking begins when the conditioning information disappears from the interpretation. A baseline calibration becomes ‘the result’, while the surrounding sensitivity and the boundary of the model are demoted to a robustness appendix. Near a threshold or singularity, however, dependence on assumptions is part of the economic result itself.

Figure 3 renders that conditionality geometrically. One line is selected from each of two illustrative equation families, and their intersection is the corresponding solution. Changing the pair changes the point before any arithmetic begins. As the pair varies, the solution moves through all four sign regions: no sign pattern follows from the unspecialised equation families alone; it belongs to a specified pair and to the modelling choices that generated it.

Figure 3. Two families of balance equations



Note: Blue and red lines: variants of the two equations; grey points: their intersections.

The hundred solutions in Section 2 were not a hundred errors. They were one point per assigned calibration from a mapping of the same logical kind, and the coefficient pair—not the arithmetic—determined what each student reported. The schematic is deliberately more general than the classroom grid. It reaches all four sign regions, whereas the particular Leontief calibration used in the prank reaches three. That difference is now informative rather than embarrassing: the figure is not another description of the classroom data but a picture of the conditional structure behind them.

The class observes the fragility of a point across neighbouring versions of a model, not several equilibria of one fixed economy. This changes the order of questions. Before asking ‘What is the equilibrium?’, the modeller should ask where the model may be interpreted, how

the solution changes across plausible parameter values, and which assumption selects the reported coordinate. A result that reverses under a small admissible perturbation deserves a different status from one that survives a broad region.

4.2. The Theorem in the Same Book

The classroom example is only a two-equation illustration. The Sonnenschein–Mantel–Debreu results make a separate and much deeper point about aggregate excess demand. Under standard conditions, aggregate excess demand retains broad formal restrictions such as continuity, homogeneity of degree zero, Walras’s law, and appropriate boundary behaviour. Within those restrictions, however, its shape can be highly general. Individual optimisation does not by itself guarantee a unique or stable market equilibrium (Debreu 1974; Mantel 1974; Sonnenschein 1972). Additional structure is needed.

Debreu’s (1970) regularity analysis answers one part of this problem: under regularity, equilibria are finite in number and locally isolated. That restores determinacy in a technical sense but does not settle the methodological question pursued here—which assumptions produced the determinacy, and what economic content do those assumptions carry?

The distinction matters. Figure 3 is neither a proof nor a special case of SMD. It shows why a solution belongs to specified equations. SMD shows that familiar restrictions on rational individuals do not supply the aggregate shape restrictions that ordinary equilibrium reasoning often presumes. The two arguments meet at one methodological question: when a determinate answer appears, which assumption produced the determinacy?

The uncomfortable feature of SMD is its institutional location. It is not a heterodox objection lodged outside the discipline. It is a theorem of general equilibrium theory, proved by general equilibrium theorists and presented within the standard graduate architecture of the subject (Mas-Colell et al. 1995, ch. 17). The same curriculum can teach the theorem and then continue with tractable equilibrium models. There is no logical contradiction in doing so: particular models may legitimately impose stronger restrictions. The blindness begins when those restrictions are treated as harmless technique rather than substantive commitments whose empirical content must be defended.

Rizvi (2006) documents how the SMD results were absorbed without reorganising the mainstream research programme, while Ackerman (2002) and Kirman (1989) draw out their implications for general equilibrium and aggregation. Mirowski’s (1989) longer history explains why equilibrium could retain authority as an organising metaphor even as its formal

content changed. The sharper accusation is not that economists have forgotten SMD. It is that a critical result can be learned, examined, and cited while leaving the reporting practice untouched. The discipline does not lack the objection; it has learned how to continue around it.

4.3. The Equilibrium Axiom as Methodological Closure

Varoufakis advances a related criticism from the side of method. Arnsperger and Varoufakis (2006) call it methodological equilibration: theoretical inquiry is organised around the question of whether behaviour approaches an equilibrium before the model has established that equilibrium is the appropriate state through which the phenomenon should be understood. Varoufakis (2014) describes the broader problem as the methodological imposition of equilibrium. In this reading, equilibrium is not always a conclusion reached after coordination has been explained; it can be the closure condition that makes a determinate analysis possible.

This should not be inflated into the claim that every equilibrium model is circular. An equilibrium concept can define a benchmark, an internally consistent set of plans, or an outcome selected by an explicit institution. The stronger criticism is more precise. Rationality alone often fails to coordinate beliefs or choose among strategically consistent outcomes. Game-theoretic refinements, conventions, expectations, learning rules, and institutional mechanisms can supply the missing selection, but each adds content that is not contained in individual rationality itself (Hargreaves Heap and Varoufakis 2004). A model may need such a rule. The methodological error occurs when the rule enters as a premise and leaves as an apparently natural prediction.

Economic indeterminacy is not the absence of rationality. It is what can remain after actors are assumed to be rational and to know that others are reasoning too, but no shared mechanism aligns their conjectures. A unique equilibrium then says more than ‘the agents are rational’. It also encodes how expectations are coordinated, how one outcome is selected, or which alternatives have been ruled out. Those additions may be defensible; they must not be invisible.

4.4. Remedies That Cost Little

The reception of the strongest criticisms might be explained by their difficulty. The neglect of weaker remedies cannot. Leibenstein (1976) did not ask economics to abandon equilibrium. His idea of inert areas, ranges within which agents do not respond to small changes, shifts attention from a coordinate towards a region in which frictions, inefficiency,

and incomplete adjustment may leave the economy. McCloskey's defence of calibration asks for less still: report the range over which the parameter values could plausibly lie and show whether the conclusion survives that range (McCloskey 2000a, 2000b). Denis (2007) identifies the underlying temptation as the hypostatisation of equilibrium—treating an analytical construction as though it were a thing discovered intact in the world.

These remedies cost little computationally. Modern software can map a region almost as easily as it evaluates a point. They cost more rhetorically. A coordinate can be placed in an abstract, compared in a table, and repeated as 'the result'. A region exposes dependence on judgement and may reveal that the preferred conclusion occupies only a narrow part of the admissible space. The persistence of point reporting is not adequately explained by technical convenience. It also preserves the authority of closure.

4.5. Knowledge Present, Knowledge Idle

The Point Equilibrium Believer is not a naive figure, and the essay would be easier but less accurate if he were. He knows that existence is not uniqueness, uniqueness is not stability, and a calibration is conditional. He may be able to state SMD on an examination and discuss equilibrium selection in a seminar. His blindness is that none of this reaches the sentence in which he reports 'the equilibrium'. Knowledge is present and methodologically idle.

That is the same structure the classroom exhibited. Had the students been asked whether a non-productive coefficient matrix or an implausibly large output change matters, they would have said yes. Nobody asked before the results collided on the blackboard. The procedure supplied a stopping rule, and knowledge that could have interrupted it remained unused. The difference between the room and the discipline is scale, not kind.

Putting the knowledge to work changes the sequence: What is the equilibrium? Under which parameter values and selection rule? Is that region economically admissible? Which mechanism could move the economy towards the point, and what evidence makes the mechanism credible? The companion simulation turns these questions into an inspection procedure. The reader moves e_1 and e_2 , crosses the Hawkins–Simon boundary, and watches the solution travel across the surfaces in Figure 2. Computation reveals the landscape. It cannot decide which part of that landscape describes the world; that judgement returns the problem to the first blind man.

The class has now produced a collection of individually defensible points and learned why no point should stand alone. The instructor's next move is to compress the disagreement into one collective verdict. The third blind man enters when the class is asked to vote.

5. The Third Blind Man: The P-Value Voter

5.1. One Sample Casts One Ballot

The class vote completes a sequence of compressions. Each student begins with a numerical solution. The instructor first replaces its magnitudes with a sign code, then combines distinct sign patterns into a policy category, and finally asks for a show of hands. The winning category is easy to announce because the information that could qualify it has already been removed. What survives is not the distribution of answers but a verdict.

A significance test can be used in the same way, although the analogy must be stated precisely. A sample S produces an estimate and a p-value $p(S)$, both of which would change under repeated sampling. A threshold α then produces a binary variable:

$$V_{\alpha}(S) = \mathbf{1}\{p(S) \leq \alpha\}, \quad \Pr_{\theta}[V_{\alpha}(S) = 1] = \pi_{\alpha}(\theta).$$

The first expression is the realised vote; the second is the probability that the procedure casts that vote under a specified state of the world. One sample therefore casts one ballot. The sampling distribution is the electorate. Treating the ballot as though it were the electorate is the statistical version of mistaking one classroom vote for the structure that generated it.

This is where the word treated matters. A significance test does not formally transform randomness into certainty. It transforms a random statistic into a decision under a chosen rule. The transformation into certainty occurs later, when ‘ $p < 0.05$ ’ becomes ‘the effect exists’, ‘ $p > 0.05$ ’ becomes ‘there is no effect’, and the sampling distribution of the verdict disappears from the prose. The formal procedure still contains uncertainty; the reported sentence has switched it off.

5.2. The Verdict Has a Distribution of Its Own

Consider a deliberately favourable benchmark. A two-sided normal test first reports $p = 0.05$. An exact replication uses the same standard error; the true effect is set equal to the first estimate; and there is no publication selection, p-hacking, or heterogeneity. Even under those assumptions, the next p-value is broadly distributed. Eighty per cent of replications lie between 0.001189 and 0.48371, and the probability of crossing 0.05 again is 0.50004. The small excess over one half is the probability of a significant result in the opposite direction. The appropriate statement is ‘approximately an even bet’, not ‘exactly one half’. This calculation belongs to the p-interval literature (Cumming 2008); the values reported here are recomputed for the two-sided normal benchmark stated above.

The calculation is not a forecast of the replication rate of the published literature. Its assumptions are intentionally kinder than reality. It isolates a logical point: a result standing on a threshold is unstable as a thresholded verdict even when the first estimate is taken at face value. The same property holds at another line. A first result at $p = 0.005$ is again approximately an even bet to pass 0.005 in an otherwise identical replication. Lowering alpha changes the type I error rate, the power, and the required sample size; it does not abolish the discontinuity created by classification. The exchange between Benjamin et al. (2018) and Lakens et al. (2018) is not resolved by pretending that thresholds are irrelevant. It shows why a threshold must be justified by the losses, design, and evidential purpose at hand rather than treated as a universal border between fact and error. Because the test is two-sided, the probability of clearing the line again is one half plus the chance of a significant result of the opposite sign: 0.5005 at $\alpha = 0.1$, 0.50004 at $\alpha = 0.05$, and 0.500000 at $\alpha \leq 0.01$ (Figure 4B).

The problem is sharper in a low-powered field. Ioannidis et al. (2017) examined 159 economics literatures containing 64,076 estimates from more than 6,700 studies and estimated median power at 18 per cent or less. Power is not the posterior probability that a published claim is true, and it is not the conditional probability that one selected result will replicate. It is the long-run probability that a design detects a specified effect when that effect is present. Yet it determines the electorate from which significant votes are drawn. In a two-sided normal test with $\alpha = 0.05$ and power 0.18, conditioning on significance alone makes the expected absolute estimate about 2.40 times the true effect. This is a type M (magnitude) error in the terminology of Gelman and Carlin (2014); 2.40 is our calculation under the stated benchmark, not a value borrowed from their examples. Selection at the line is sufficient to generate large exaggeration before any additional bias is introduced.

5.3. The Same Compression at Two Scales

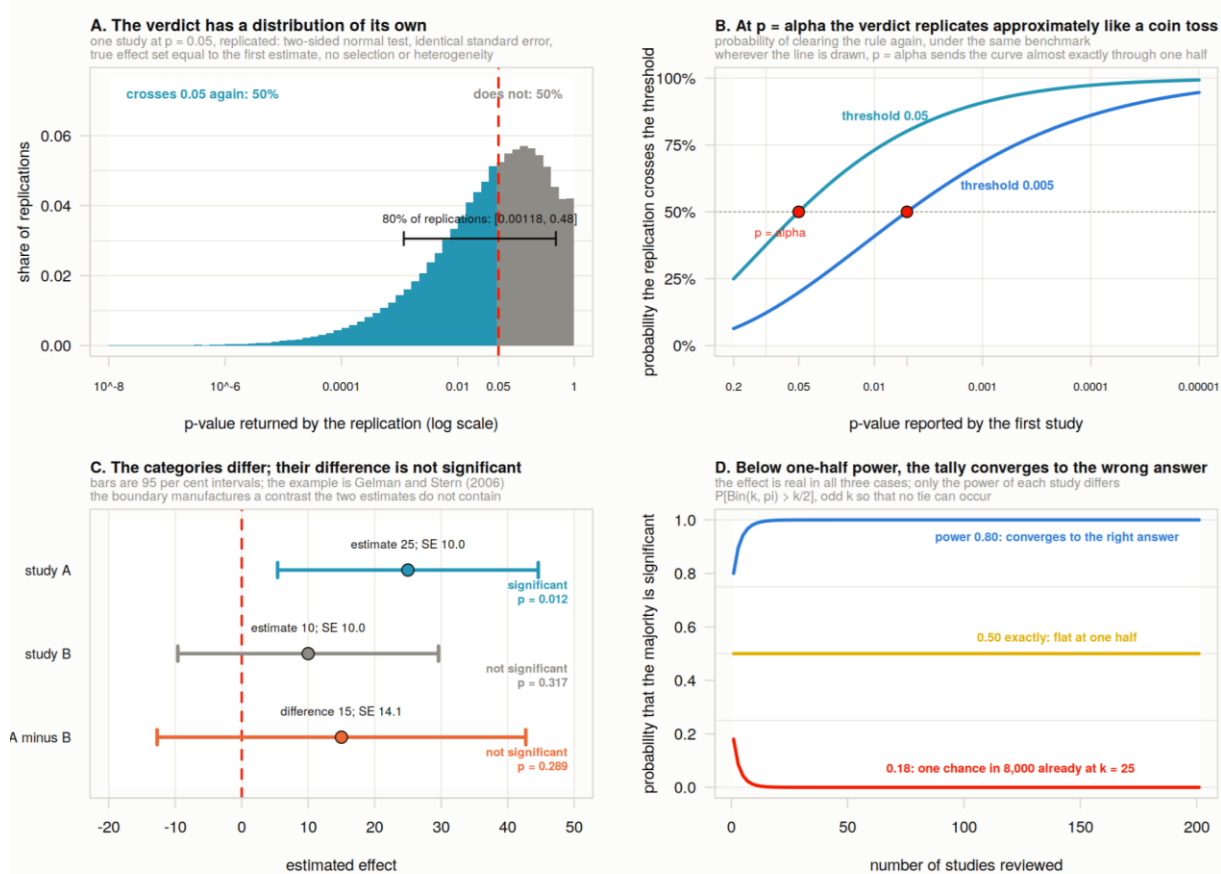
At the scale of two studies, Gelman and Stern (2006) show what the boundary manufactures. One estimate is 25 with a standard error of 10 and is called significant; another is 10 with the same standard error and is called not significant. Their difference is 15 with a standard error of 14.1 and is itself not significant. The categories differ, but the estimates have not been shown to differ. A continuous comparison has been replaced by two labels, and the labels are then compared as if they were measurements.

At the scale of a literature, vote counting repeats the operation. If k independent studies each have power π against a real effect and k is odd, the probability that a majority is significant is

$$\Pr\{\text{majority significant}\} = \Pr\{\text{Bin}(k, \pi) > k/2\}.$$

When π exceeds one half, the tally converges to the right answer as the literature grows. At $\pi = 0.50$ it remains exactly one half for every odd k . Below one half it converges to the wrong answer: most studies fail to cross the line even though the effect is real in every case. At power 0.18, the probability that a majority of twenty-five studies is significant is about one in 8,000. More voters make the categorical conclusion more decisive and less correct. This is the paradox identified by Hedges and Olkin (1980): vote counting can lose power as the number of studies reviewed increases.

Figure 4. The p-value as a ballot: one replication and many studies



Source: Own calculations; Panel C based on Gelman and Stern (2006), Panel D based on Hedges and Olkin (1980).

Note: A: replication distribution of p after a first $p = 0.05$ (benchmark of Section 5.2). B: probability of clearing the threshold again. C: estimates (standard errors) with 95 per cent intervals. D: probability that a majority of k studies is significant at power π .

The realised verdict loses information; counting realised verdicts does not restore it.

A single classification removes uncertainty from view. A stack of classifications can lend that loss the appearance of collective certainty. Panels C and D are one argument at two scales: what the boundary does to a comparison between two studies, it does again—and irreversibly—to a reading of an entire literature. Gelman and Stern expose the local category error; Hedges and Olkin show what happens when the category error is aggregated.

5.4. What the Vote Can and Cannot Decide

McCloskey (1998) criticises a routine in which empirical inquiry is organised around formulating a hypothesis, fitting a model, testing for statistical significance, and publishing the resulting verdict. The routine is attractive because it supplies a stopping rule that appears impersonal. But a p-value answers a limited conditional question: how incompatible is the observed statistic, or something more extreme, with a specified statistical model? It is not the probability that the null hypothesis is true, not the magnitude of the effect, not its economic importance, and not a measure of the cost of being wrong (Wasserstein and Lazar 2016). Each of those questions requires information that the vote does not contain.

The response should not be an anti-p-value manifesto. A pre-specified threshold can be useful when a real decision must be made and its error costs, design, and multiplicity are explicit. The same number becomes misleading when a convenient convention silently substitutes for those judgements. Amrhein et al. (2019) object not to calculation but to dichotomous declarations that erase compatibility ranges, magnitude, and context. The minimal reform is demanding: report estimates and uncertainty, justify any decision rule, distinguish absence of evidence from evidence of absence, and do not infer a difference from different labels.

A larger sample can reduce sampling uncertainty and often should. It cannot decide whether the specification identifies the intended causal relation, whether heterogeneous agents can be aggregated into the object being estimated, or whether a stable point describes a nonlinear dynamic system. Those are not imperfections in the electorate. They change the question on the ballot.

5.5. What Is Being Voted On?

The classroom's 57–43 division and the 18 per cent median power reported for economics must not be treated as the same quantity. The first is a deterministic share of coefficient assignments under the prank; the second is a repeated-sampling property of empirical designs under specified effects. A bar chart placing them side by side would manufacture a comparison.

The legitimate contrast is institutional: nobody would call a 57–43 classroom vote unanimous, yet empirical prose can become categorical even when the procedure that produced it would return the same category much less reliably.

The next three blind men widen the electorate in a different sense. Section 6 asks how leverage, misspecification, and outliers change the estimate before the test is run. Section 7 asks whether individual relations survive aggregation. Section 8 asks whether a correctly derived transition rule converges at all. They do not explain away sampling variation; they reveal that repeated samples may be voting on a fragile, aggregated, or dynamically misleading object. More ballots cannot repair a ballot whose question has not been defended.

The P-Value Voter's stopping rule is not merely $p < 0.05$. It is the belief that classification completes inquiry. If every sample-dependent estimate must be reported with an account of its uncertainty, why is the binary verdict produced from those estimates so often exempted from the same requirement?

6. The Fourth Blind Man: The Econometrician with Perfect Data

The third blind man asks whether an estimated coefficient crosses a threshold. The fourth asks what makes that coefficient an estimate of the economic relation we care about. A vast dataset can make a coefficient extremely precise while leaving the design unable to separate its causes—or while precisely transmitting a violation of the relation into the estimate. Perfect measurement is not perfect identification.

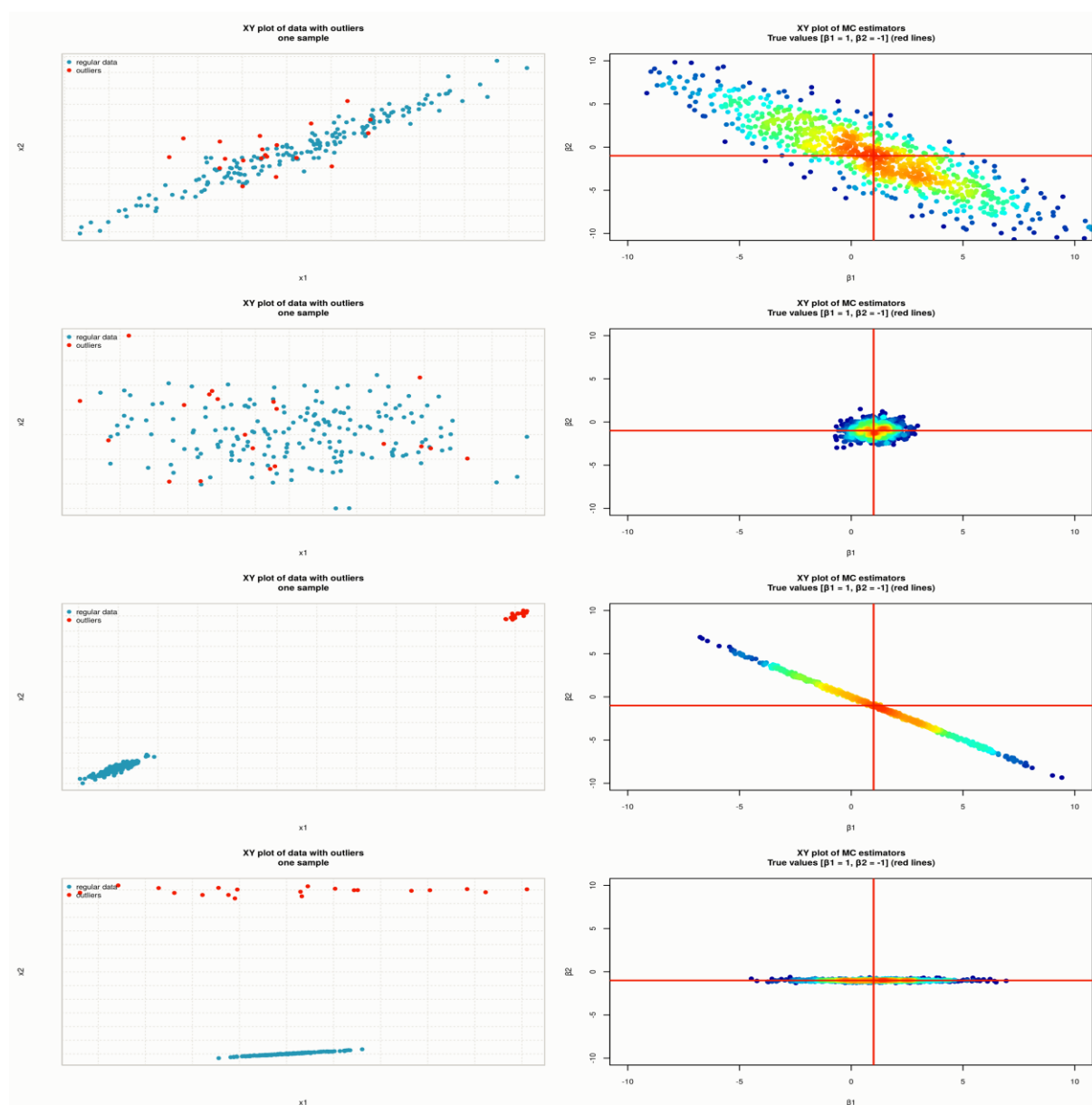
The classroom prank already contains this problem. The input–output coefficients are measured or calibrated, and the policy response is obtained through the inverse $(I - A)^{-1}$. Regression uses a different model but the same dangerous operation: $\hat{\beta} = (X'X)^{-1}X'y$. In both cases inversion turns weak separation in the underlying system into large movements in the reported answer. The mathematics is shared; the economic meaning of failure is not.

6.1. Four Designs, One Correctly Specified Model

The simulation begins with $y = a + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ and true slopes $\beta_1 = 1$ and $\beta_2 = -1$. Each design has 180 regular and twenty unusual observations ($n = 200$, $s = 0.10$), with independent zero-mean normal errors of standard deviation $\sigma = 5$. Figure 5 holds the data-generating relation correct and changes only the x_1 - x_2 geometry. The estimator clouds pose the puzzle: sample size and model stay fixed, yet information about the two slopes changes radically. The values reported in this section are stable across random seeds; the companion dashboard lets the reader change the seed.

From top to bottom, Figure 5 shows a near-collinear cloud with scattered unusual observations, an approximately orthogonal design, a distant cluster aligned with the cloud, and a cluster displaced mainly in x_2 . Every regression is correctly specified, yet precision and dependence differ sharply. Given as regular cloud (standard deviation of x_1 and x_2 , correlation) and unusual group (standard deviation, correlation, centre), the parameters are: (a) 0.20, 0.97; 0.18, 0, (0, 0); (b) 0.50, 0; 0.70, 0, (0, 0); (c) 0.30, 0.90; 0.20, 0.90, (10, 10); (d) 0.12, 0.99; (0.50, 0.25), 0, (0, 12). The regular cloud is centred at the origin, the intercept is $a = 10$, and each estimator cloud contains 900 replications.

Figure 5. Four regressor geometries and their estimator clouds



Note: Left: one x_1 – x_2 sample (teal: 180 regular, red: 20 unusual observations). Right: (β_1, β_2) over 900 replications; red lines: true values $\beta_1 = 1, \beta_2 = -1$. Rows (a)–(d) as described in Section 6.1.

6.2. Inversion, Direction, and Information

The Hawkins–Simon condition is easiest to understand as a requirement that production can reproduce itself before it satisfies final demand. An extra unit of final demand requires direct output; producing that output requires intermediate inputs; producing those inputs requires further inputs, and so on. In a productive economy these successive rounds shrink, so the total requirement is finite and non-negative. For the classroom matrix, the first leading principal minor, $1 - a_{11} = 0.40$, is fixed and positive; productivity therefore depends on the second, $D = \det(I - A)$, also being positive. When D approaches zero, the rounds cease to die out and the Leontief multiplier explodes. When D is negative, the equations may still yield an exact pair of numbers, but those numbers no longer describe a productive economy capable of replacing its intermediate inputs and meeting final demand. This is the first paradox of the exercise: algebraic solvability survives after economic admissibility has failed.

In regression, the analogous warning concerns the covariance matrix Σ of the regressors. With two regressors of variances V_1 and V_2 and correlation r , $\det(\Sigma) = V_1 V_2 (1 - r^2)$. As $|r|$ approaches one, the data contain abundant movement but little independent movement: x_1 and x_2 travel almost as a bundle, so the model cannot reliably decide which coefficient produced the observed change in y . The fitted values may remain stable even while the individual slopes become fragile. For the two slope estimates, let X_c denote the full-rank matrix of regressors centred on their sample means. Assume conditionally mean-zero, homoskedastic and uncorrelated errors. Conditional on this design,

$$\text{Var}(\hat{\beta}_s | X_c) = \sigma^2 (X_c^T X_c)^{-1} \approx \frac{\sigma^2}{n} \Sigma^{-1}.$$

$$\text{Corr}(\hat{\beta}_1, \hat{\beta}_2 | X_c) = -r_X.$$

Here $\hat{\beta}_s$ is the vector of the two slope estimates and r_X is the correlation of the regressors in the realised sample. The approximation uses their population covariance matrix Σ ; the conditional correlation identity is exact. Near $\det(\Sigma) = 0$, the inverse stretches sampling noise most strongly in the direction the data distinguish least well. This resembles the classroom boundary: $\det(I - A)$ near zero amplifies uncertainty about technological requirements, whereas $\det(\Sigma)$ near zero amplifies uncertainty about separate statistical effects. One threatens the admissibility of an economy; the other threatens the attribution of an effect. In both, a clean calculation can conceal that the answer depends on a nearly missing distinction.

The top design in Figure 5 has an estimator correlation of approximately -0.90 and a standard deviation of β_1 of about 4.0. The approximately orthogonal design reduces that standard deviation to about 0.68. Moving the unusual observations far away does not settle the matter. A cluster aligned with the correlated cloud makes the two estimates almost perfectly dependent (correlation approximately -0.999), whereas displacement mainly in x_2 yields a standard deviation of about 1.9 for β_1 but only 0.10 for β_2 . The same twenty points can clarify one coefficient, leave another weak, or bind the two estimates into a trade-off.

The pooled covariance matrix makes the mechanism visible. Let s be the share of the unusual group, Σ^R and Σ^O the within-group covariance matrices, and δ the displacement of the unusual group's centre relative to the regular group's centre. Then

$$\Sigma = (1 - s)\Sigma^R + s\Sigma^O + s(1 - s)\delta\delta'$$

For a non-zero displacement, the final term has rank one. Moving a cluster adds information mainly along the direction δ ; it does not improve the design equally in every direction. This is why direction can matter more than distance. For the first geometry, with $n = 200$ and $s = 0.10$, the covariance approximation gives a standard deviation of β_1 of about 3.81 when the group is moved six units along the existing cloud and about 0.20 when it is moved the same distance along x_1 , a nineteen-fold difference. Replacing twenty observations in a cloud with correlation 0.97 by approximately uncorrelated observations at the same centre reduces the approximate standard deviation of β_1 from 7.27 to 3.91. Along the existing direction of collinearity, by contrast, the limiting gain relative to the centred mixture is only about 2.5 per cent when both the regular cloud and the displaced cluster have within-group correlation 0.90, as in the third design.

This creates the first paradox of the regression experiment. Observations that look like outliers can improve precision, while observations lying naturally along the dominant cloud can add almost no identifying information. Distance is not a verdict. The companion dashboard maps the full displacement plane; the result retained here is the economic lesson: data are valuable when they supply the variation required to distinguish the mechanisms named by the model.

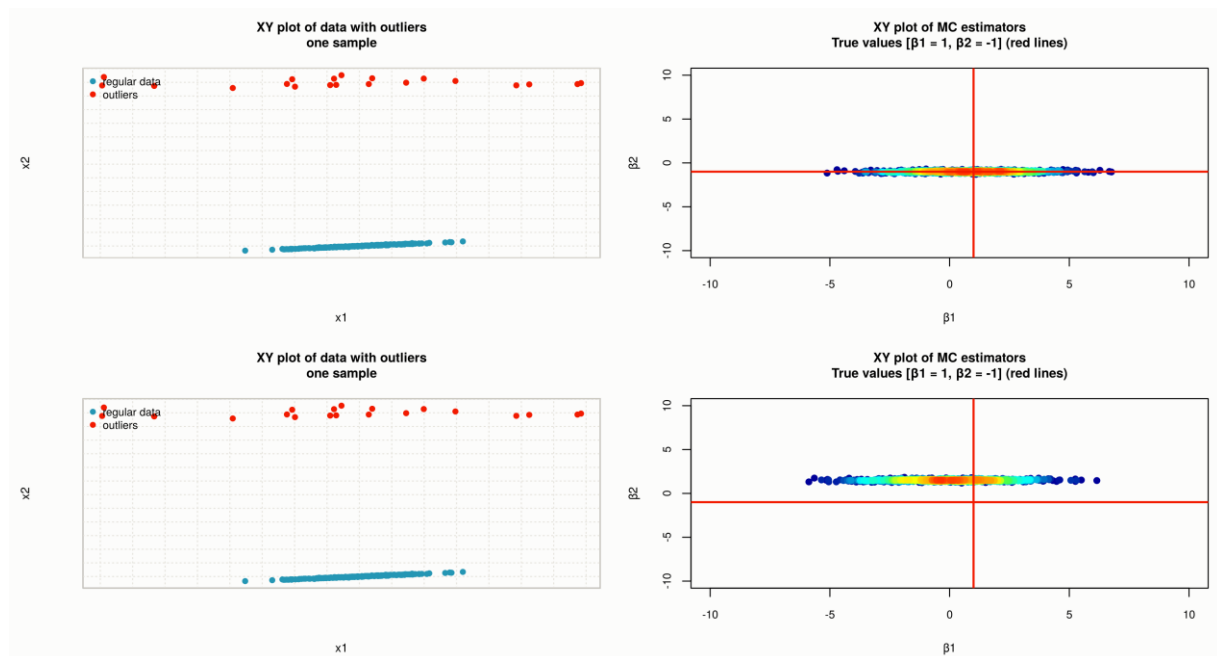
6.3. The Same Leverage, a Different Relation

A design can separate the regressors and still estimate the wrong relation. Figure 6 holds every x_1 and x_2 value fixed. In its top row, the twenty high-leverage observations obey the regression plane. In the bottom row, the same observations are lifted by 30 units in y .

The two regressor plots are identical: nothing in the x_1 - x_2 geometry reveals whether the unusual points support or violate the conditional relation. Only the residual dimension carries that information. The estimator cloud remains narrow but moves to the opposite side of the true value of β_2 : regressor leverage reveals influence, not validity.

The distinction between leverage—the regressor geometry that makes an observation capable of moving a fit—and influence—the realised change in the fitted result—is standard in regression diagnostics (Belsley et al. 1980; Cook 1977). What the experiment adds is the economic question that follows the diagnostic: does the influential observation identify the relation of interest, or efficiently transmit a violation of it?

Figure 6. The same high-leverage observations, obeying and violating the relation



Note: Top: unusual observations on the true regression plane. Bottom: the same observations shifted by +30 in y . Left: identical x_1 - x_2 designs; right: estimator clouds.

At population level, the group-shift construction gives a compact expression for the asymptotic slope bias. Write $y = a + x^\top \beta_s + cG + \varepsilon$, where G identifies the unusual group, $s = P(G = 1)$, and $\delta = E(x | G = 1) - E(x | G = 0)$. For independent within-group samples with limiting group share s , assume $E(\varepsilon | x, G) = 0$, finite second moments, and nonsingular $\Sigma = \text{Var}(x)$. The estimated slopes then converge to the population linear-projection coefficient, whose displacement from the true slope vector β_s is

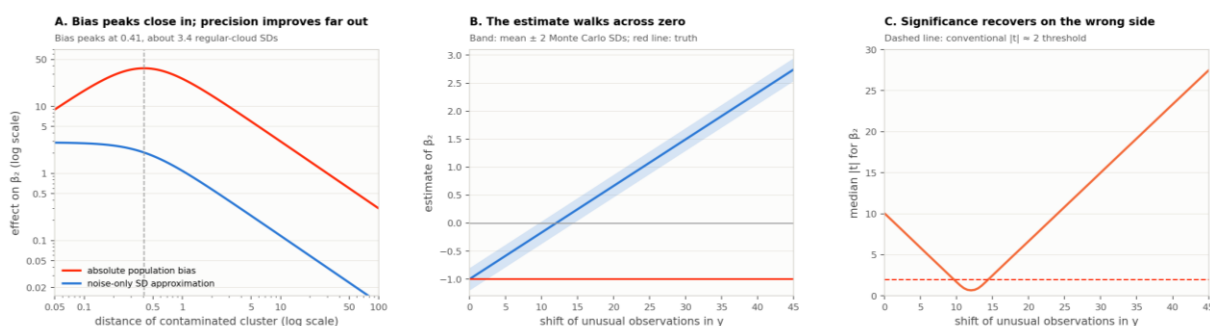
$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_s - \beta_s = cs(1-s)\Sigma^{-1}\delta.$$

The same inverse that governs precision now transmits a systematic violation. When $c = 0$, leverage can improve precision without moving the centre of the estimator distribution. When $c \neq 0$, the same leverage can transmit the violation with equal efficiency. At a vertical shift of 30, the first two geometries in Figure 5 have zero population slope bias because their group centres coincide, $\delta = 0$. In the two displaced-cluster geometries, the Euclidean magnitudes of the population slope-bias vectors are approximately 2.10 and 2.64. Small non-zero Monte Carlo averages in the centred designs are simulation error, not evidence of population slope bias. A small standard error can mean either that the design has isolated the intended relation or that it has isolated a contaminated relation with exceptional efficiency.

Nor is the most distant cluster necessarily the most damaging. Figure 7A moves the contaminated group away from the regular cloud. Population bias first rises, peaks at a distance of about 0.41 (roughly 3.4 marginal standard deviations of the regular cloud) and then declines, while the noise-only approximation to the standard deviation continues to fall. Moderately unusual observations can thus create more bias than spectacular outliers because they combine enough leverage to move the estimate with enough overlap to remain integrated into the fitted relation.

6.4. From Truth, Through Insignificance, to Significant Falsehood

Figure 7. Contamination in y and significance on the wrong side



Note: A: absolute population bias of β_2 and its noise-only standard deviation against displacement (vertical shift 30). B: mean $\beta_2 \pm 2$ Monte Carlo standard deviations; red line: $\beta_2 = -1$. C: median $|t|$.

The numerical path in Figures 7B and 7C links the Econometrician with Perfect Data directly to the P-Value Voter. With true $\beta_2 = -1$ and no vertical contamination, the mean estimate is -1.00 , its standard deviation is 0.100 , and the median $|t|$ is 10.0 . At a vertical shift of 12 , the mean estimate approaches zero and the median $|t|$ falls to about 0.7 . At a shift of 30 , the mean estimate is approximately $+1.50$ while its standard deviation remains about 0.101 and median $|t|$ rises to 15.0 . At 45 , median $|t|$ reaches 27.5 . Between shifts of zero and 30 ,

median R^2 rises from about 0.35 to 0.54. The wrong-signed coefficient is similarly precise, more significant by the conventional test, and embedded in a better-fitting regression than the coefficient centred on the truth. These panels hold the displacement at 12 in x_2 and vary the vertical shift, with 200 observations per sample (twenty of them unusual), the remaining parameters of geometry (d) in Figure 5, 12,000 Monte Carlo replications per point, and common regressor and error draws across shifts. The companion dashboard reproduces them on the page ‘The Econometrician’, tab ‘In the article’.

This is the second and strongest paradox of the regression experiment. The path from truth to falsehood passes through insignificance. A threshold first suspends the verdict and then restores an even more decisive verdict on the wrong side. The fitted regression reports a small p-value around a precisely estimated but contaminated coefficient. The conventional t-test is not thereby guaranteed to retain its nominal size under misspecification, and its reported significance does not establish that the specification still represents the economic relation that gave the coefficient its meaning. Stability was never a property of the answer; it was a property of the rule that produced it.

6.5. Precision Is Not Identification

The wider family includes omitted variables, measurement error, sample selection, simultaneity, weak instruments, unstable populations, functional-form misspecification, and extrapolation beyond the support of the data. Causality is an especially revealing case. A theoretical model may specify the mechanism while an estimated equation retains only some variables, timing assumptions, and ceteris paribus conditions. The coefficient can then inherit causal language from the theory without carrying the design that would identify the causal claim.

A terabyte of administrative records can reduce sampling uncertainty sharply. It cannot determine which observations obey the same relation, which direction of variation identifies the target, whether the functional form is stable, or whether the coefficient travels to the policy setting. The Econometrician with Perfect Data becomes blind when precision substitutes for identification, specification, and interpretation. Even after those problems have been confronted, one question remains: what permits relations estimated for heterogeneous individuals to describe the economy as a whole? That is the problem taken up by the fifth blind man.

7. The Fifth Blind Man: The Axiomatist

After the reveal, some students ask the question the exercise was built to provoke: why should there be one group answer at all? Each student solved a different economy. The instructor nevertheless requested one recommendation, and the class supplied one. The prank turns aggregation from a routine operation into something the room has already done to itself.

7.1. When the Whole Looks Like the Part

Begin with an accommodating version of the exercise. Give every student the same productive coefficient matrix. Everyone solves the same system, obtains the same point, and every summary of the class coincides: the mean, the median, the majority category, and the solution of a representative student are one number. Aggregation is effortless because heterogeneity has been removed before the analysis begins.

The prank restores the variation. Student i solves $z_i = z(e_{1i}, e_{2i})$, where $z(e) = (I - A(e))^{-1} \Delta d$. The class can now be described by several distinct objects: the solution of a model built from average coefficients, the mean and median of the individual solutions, the four exact sign patterns, and the three compressed policy categories.

The familiar caution is that these objects need not coincide. The system is linear in Δx , while its solution depends nonlinearly on the coefficients through matrix inversion:

$$z(\bar{e}_1, \bar{e}_2) \neq \frac{1}{N} \sum_i z(e_{1i}, e_{2i})$$

In a well-conditioned problem, this is a mild non-commutativity. If the coefficients are drawn uniformly from a narrow square around the baseline, with half-width $h = 0.004$, the representative solution for Δx_1 is 100.0, the population median is 100.0, and the population mean is approximately 105.1. The summaries differ by about five per cent, but all are finite and stable. The classroom exercise does not operate in that range.

The gentle Jensen-type discrepancy is only the benchmark. Once the parameter range reaches the singularity, aggregation changes in kind: one of the population summaries ceases to exist as a finite moment.

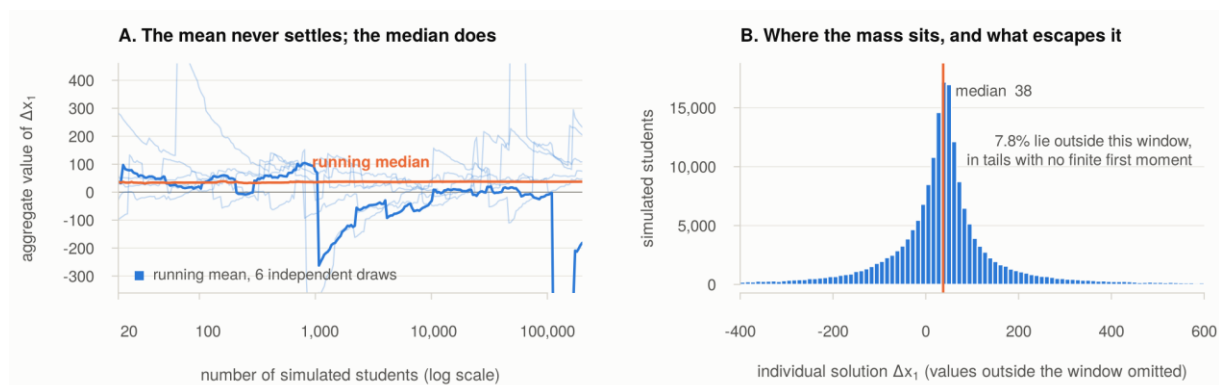
7.2. A Threshold, and One Summary That Ceases to Exist

Every individual solution is a ratio whose denominator is the determinant D . For the classroom calibration, $D(e_1, e_2) = 0.006 - 0.54e_1 - 0.40e_2$. If the two contested coefficients are drawn from the symmetric square $[-h, h]^2$, the smallest attainable determinant is $0.006 - 0.94h$. The population first touches the singularity at

$$h^* = \frac{0.006}{0.94} \approx 0.00638$$

The continuous simulation of the classroom exercise uses $h = 0.05$, almost eight times this threshold. The determinant then changes sign inside the parameter square. Because its density is positive at zero, the solution has Cauchy-type tails: its density decays as the inverse square, its first absolute moment is infinite, and its ordinary population mean is undefined. The problem is not that the mean is difficult to estimate. There is no finite population value for a sample mean to estimate. Appendix A gives the derivation.

Figure 8. Aggregating individual solutions near the Hawkins–Simon boundary



Note: A: running means of Δx_1 in six simulated populations (blue) and the running median in the highlighted one (orange). B: central mass of the individual solutions. $h = 0.05$.

Across six populations of 200,000 simulated students (seed 5 in the companion dashboard), the final medians lie between 37.53 and 37.77. The corresponding running means end between -181.0 and $+232.3$, disagree in sign, and remain vulnerable to the next draw. Another seed gives other means, often of another order of magnitude, while the medians stay close to 37.6; the instability of the mean is the result, not noise around it. The vertical jumps in Panel A mark the arrival of a single economy with D close to zero; one locally assigned model is then able to move the average of the entire simulated class. The vertical axis of Panel A is limited to -330 to 430 , so the largest excursions of the means leave the window; in Panel B about 7.8 per cent of the individual solutions lie outside the plotted range, in tails with no finite first moment.

The representative model is insulated from this change because it depends only on the centre of the coefficient distribution. It returns $\Delta x_1 = 100.00$ both below and above h^* . The population it claims to represent has changed completely: its median falls from 100.0 to approximately 37.6, and 38.9 per cent of its matrices violate Hawkins–Simon. That share belongs to the continuous population, not to the hundred assigned pairs, and the population mean no longer exists as a finite moment. The representative model does not gradually become a poorer approximation. It continues to return the same confident number after the object identified as the population mean has disappeared.

7.3. What Survives Aggregation, and Why That Is Not Reassuring

The arithmetic mean fails because it gives unbounded leverage to the tails. The median survives because a quantile depends on rank rather than magnitude. The majority sign also converges: it averages bounded indicator variables, so isolated explosions in Δx_1 cannot dominate it. The surprising ranking is that the summary using every numerical magnitude is the one without a population limit, while the coarser summaries remain stable.

Stability still does not confer economic validity. The median of approximately 37.6 summarises a population in which 38.9 per cent of the coefficient matrices fail the Hawkins–Simon condition. It is a stable centre of a mixture of productive and non-productive economies. The majority sign discards even more: magnitudes, distances from the singularity, and admissibility all vanish behind one category.

The same mechanism now connects three blind men. In Sections 2 and 5, a stable vote included economically impossible assignments. In Section 6, a precise t-statistic attached itself to a coefficient with the wrong sign. Here, a stable median describes a population containing inadmissible models, while the arithmetic mean has no finite target. In all three cases, the well-behaved statistic is not the one that answers the economic question, and its good behaviour is what makes it persuasive.

The dashboard turns the threshold into a control. Moving h below and above h^* switches the heavy-tail phenomenon off and on while leaving the central calibration unchanged. This isolates what follows from nonlinear matrix inversion from what follows when the population crosses the admissibility boundary.

7.4. What the Axiomatist Contributes

The Axiomatist asks the indispensable question: under exactly which conditions does a property established for individuals survive aggregation? Gorman's (1953) aggregation result gives precise content to the phrase 'a particularly accommodating structure'. Exact representation requires strong restrictions on individual preferences and income effects, commonly expressed through a shared affine structure of Engel curves. Homogeneity, common parameters, separability, and additive interactions play analogous roles in other aggregation problems: they allow a heterogeneous population to behave like a scaled-up individual and determine which forms of heterogeneity disappear before the derivation begins.

Kirman (1992) drew the methodological conclusion. A representative individual is not a harmless label for the aggregate: in general, there is no reason for the aggregate to preserve the properties imposed on its members, or for the representative agent's ranking of policies to coincide with the rankings of the people represented. The classroom result provides a sharp numerical prototype. Averaging coefficients, averaging solved models, taking a median, and voting over signs are different operations, with different existence conditions and different economic meanings.

The aggregate side of the Sonnenschein–Mantel–Debreu result was discussed in Section 4. Its role here is narrower: it shows why regularity at the level of individual choice does not automatically become regularity of aggregate excess demand (Mas-Colell et al. 1995). The classroom analysis identifies a different aggregation failure—a nonlinear map from coefficients to solutions and, after h^* , heavy tails that destroy a particular population moment. Together with Gorman's and Kirman's results, the theorem reinforces the same warning: every bridge from a well-behaved agent to a well-behaved economy requires assumptions of its own and a defence of what they exclude.

Blindness begins when the conditions enabling aggregation are read as a description of the world rather than as a choice about which world is being modelled.

7.5. Aggregate First or Solve First?

The companion dashboard makes this choice operable. At $h = 0$ every simulated student receives the same parameters and all summaries coincide. As h rises, the coefficient pairs spread around a fixed centre, isolating heterogeneity from any change in the representative model. The dashboard reports the representative solution, sample mean, median, majority category, full solution cloud, productive share, and whether the current range crosses $D = 0$.

The finite classroom grid and the continuous population answer different questions. Every individual solution and the mean and median of the one hundred assigned solutions are finite; claims about what happened in the room belong to that finite design. The continuous population is the correct object for asking whether a population moment exists. Reporting its sample mean without stating that the corresponding population mean is undefined would repeat, in the analysis of the prank, the substitution the prank was designed to expose.

Restoring heterogeneity opens the next step. Interaction, feedback, and time can make the aggregate itself evolve, and the sixth blind man begins where a law of motion is added.

8. The Sixth Blind Man: The Chaos Hunter

The first five blind men fail in the same direction. Each trusts an instrument that works and lets its successful output stand in for the economic object: a number for interpretation, a point for a landscape, a threshold for evidence, precision for identification, or an aggregate structure for the people it represents. Their blindness is an excess of commitment to something they possess.

8.1. A Blindness That Works by Denial

The sixth is different. The Chaos Hunter does not merely extend his own tool too far; he rejects someone else's. He has watched equilibrium analysis remove time, feedback, path dependence, heterogeneous interaction, and emergence. He approaches the model with a prior verdict: equilibrium is not a partial description but a mistaken concept, and the task is to demonstrate its failure. The rebellion has a real target. Its blindness lies in turning that target into a negation of the whole tradition.

The mirror image is the Point Equilibrium Believer. One computes a fixed point and stops; the other produces a complicated trajectory and treats the fixed point as discredited. One touches the point, the other the path. Neither sees that both belong to the same law of motion. The sixth blind man comes last because this posture threatens the remedy proposed in Section 9: a specialist who denies that the others are describing anything has removed the common object around which conversation can begin.

8.2. Instability Is Not Yet Chaos

The classroom prank is a static linear system. Near the Hawkins–Simon boundary, a small change in the coefficients can produce a large change in the solution, reverse its sign, or move

the assigned economy outside the productive domain. That is the sensitivity of an algebraic system. It is dramatic, economically consequential, and still not deterministic chaos.

Adding time does not by itself change the category. A classical dynamic input–output system remains linear: depending on its adjustment structure, it can converge, oscillate, or diverge, and its price and quantity dynamics can exhibit the dual stability problem analysed by Jorgenson (1960, 1961). Static sensitivity, dynamic instability, a periodic cycle, and deterministic chaos are different claims. Sensitivity compares solutions across nearby parameters. Instability means that deviations grow, often without bound. A cycle repeats a finite sequence. Chaos remains bounded and aperiodic while nearby trajectories separate exponentially.

Chaos requires two opposing operations. Stretching drives nearby trajectories apart; folding returns them to a bounded region. An explosive linear model stretches without folding and therefore diverges instead of becoming chaotic. The classroom saying ‘add time, nonlinearity, and a third body’ is a useful mnemonic, but the third body is an image rather than a theorem. A smooth autonomous continuous-time system needs at least three state variables to exhibit chaos, as in Lorenz (1963); a discrete-time system can become chaotic with one state variable and one nonlinear rule, as May (1976) demonstrated. What matters is a mechanism that both stretches and folds.

8.3. Chaotic Results Are Cheap; Explanatory Bridges Are Expensive

Once this distinction is visible, it becomes easy to manufacture chaos. Add a period index, bend an equation, introduce a delay or threshold, and search the parameter space until the largest Lyapunov exponent becomes positive. The attractor is mathematically genuine, but its existence establishes only what the altered model can generate. It does not establish that the added feedback belongs to an economy, that the required parameter region is credible, or that the mechanism survives a change in specification.

The positive Lyapunov exponent can become the p-value of nonlinear dynamics. A classification threshold replaces a substantive question: $p < 0.05$ becomes ‘an effect exists’, while $\lambda_{\max} > 0$ becomes ‘the economy is chaotic’. In both cases, the binary label compresses magnitude, mechanism, admissibility, and robustness. The important questions begin after the label appears. Which equation had to bend? How far? What economic restriction produced the fold? What observation would distinguish the proposed deterministic mechanism from noise or structural change?

Chaotic construction is cheap because parameter search can supply examples. A bridge is expensive because it must start from assumptions economists already use, preserve their economic meaning, derive the law of motion from an explicit problem, and still produce the dynamics the Chaos Hunter wants to reveal. Nishimura and Yano's two-sector optimal-growth model is valuable precisely because it pays that price.

8.4. Equilibrium inside Chaos: A Minimal Leontief Bridge

Nishimura and Yano (1995) consider a two-sector economy in which the consumption and investment goods are produced with fixed-proportions technologies. In the normalisation used by Nishimura and Stachurski (2004),

$$F^C(K^C, L^C) = \min\{K^C, L^C\}, \quad F^K(K^K, L^K) = \lambda \min\{K^K, L^K/b\}.$$

The utility function is linear, capital depreciates fully in one period, and $b > 1$ makes the investment-good sector more labour-intensive. Let $\gamma = b - 1$. For suitable parameters, the solution to the intertemporal optimisation problem is the single-valued transition map

$$h(k) = \lambda k, \quad 0 \leq k \leq \frac{1}{\gamma + 1},$$

$$h(k) = -\frac{\lambda}{\gamma}(k - 1), \quad \frac{1}{\gamma + 1} \leq k \leq 1.$$

The two branches are straight and meet at a kink. The map is an economically derived tent map: the nonlinearity is not imported from physics but arises when the binding factor changes in a fixed-proportions technology. A transparent set of sufficient conditions makes this map both the optimal policy and expansive:

$$0 < \rho < 1, \quad \gamma > 0, \quad \rho\lambda > 1, \quad \gamma + 1 > \lambda,$$

$$\gamma < \lambda \leq \min\left\{\frac{\gamma + \sqrt{\gamma^2 + 4\gamma}}{2}, \frac{-1 + \sqrt{1 + 4\gamma}}{2\rho}\right\}.$$

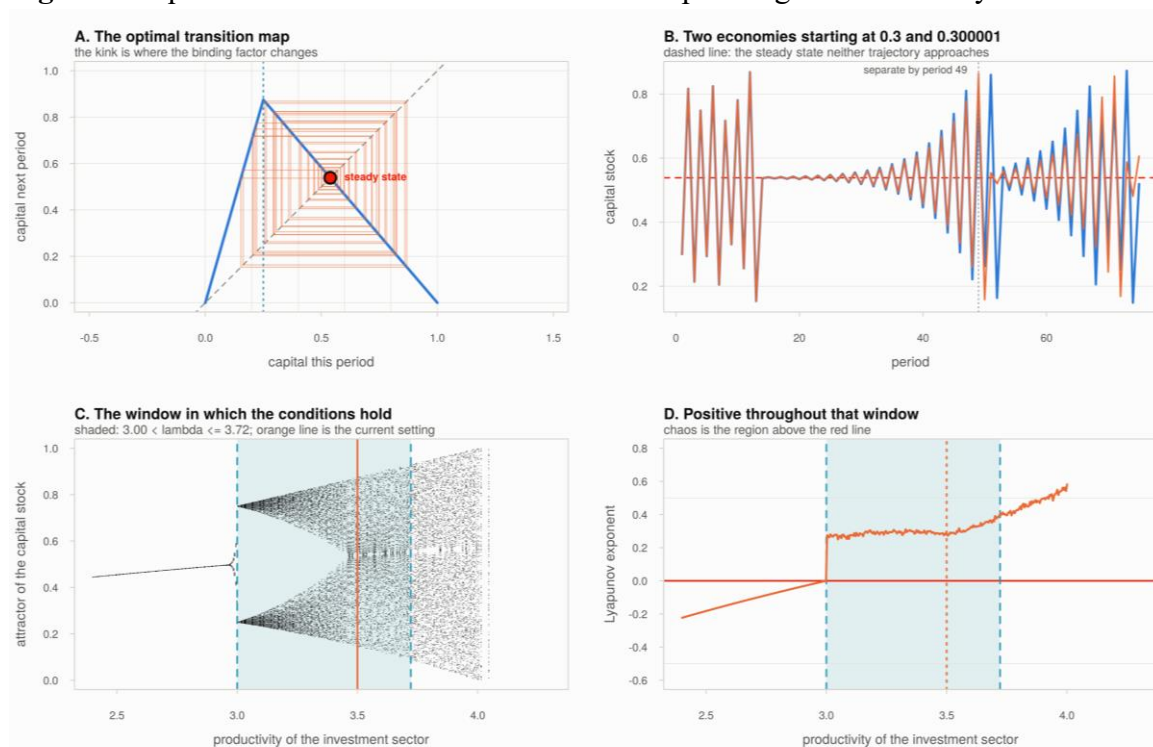
Figure 9 uses $b = 4$, so that $\gamma = 3$, together with $\lambda = 3.5$ and $\rho = 0.35$. The checks are explicit: $\rho\lambda = 1.225 > 1$, $\gamma + 1 = 4 > 3.5$, and the upper bound is 3.722. The displayed parameterisation therefore lies inside the sufficient optimal-chaos window $3 < \lambda \leq 3.722$.

The mechanism is visible in the slopes. The rising branch has slope λ and the falling branch has slope $-\frac{\lambda}{\gamma}$. With the displayed values, these are 3.5 and -1.17 . Neither λ nor b is a share, and neither is constrained to lie below one. Stretching is generated jointly

by investment productivity and the relative labour requirements of the two technologies. Folding comes from the kink at which the binding factor changes.

The steady state is $k^* = \frac{\lambda}{\gamma + \lambda} = 0.5385$. It is correctly computed, lies on the falling branch, and is repelling because the local slope is $-\frac{7}{6}$. The chaotic attractor remains inside the strictly positive interval $\left[\frac{7}{48}, \frac{7}{8}\right]$. Two economies starting from initial capital stocks $k_0 = 0.3$ and 0.300001 separate visibly after about forty-nine periods, and the largest Lyapunov exponent at the displayed parameterisation is approximately 0.280. Figure 9 shows the optimal transition map with a seventy-iteration cobweb (Panel A), the two separating trajectories (Panel B), the attractor as λ varies (Panel C), and a positive largest Lyapunov exponent throughout the shaded window (Panel D).

Figure 9. Equilibrium inside chaos in a two-sector optimal-growth economy



Source: Own calculations based on Nishimura and Yano (1995) and Nishimura and Stachurski (2004, Section 4.3).

Note: $b = 4$ ($\gamma = 3$), $\lambda = 3.5$, $\rho = 0.35$. A: optimal transition map and cobweb; B: trajectories from $k_0 = 0.3$ and $0.3 + 10^{-6}$; C: attractor against λ , shaded window $3 < \lambda \leq 3.722$; D: largest Lyapunov exponent.

The discount factor needs two statements, not one. The displayed example uses $\rho = 0.35$ because it satisfies the simple sufficient conditions above; this particular construction has a non-empty expansive region when $\rho < 1/2$. Those conditions do not define the general limit of optimal chaos. An alternative construction is discussed in the original paper and in Nishimura and Stachurski (2004), extending the compatibility of optimal chaos with patient discounting

beyond the transparent sufficient conditions used here. The present article does not rely on that stronger result: Figure 9 is interpreted only within the explicitly checked parameter window.

This is not the same matrix input–output model used in the prank. It is a two-sector growth model built with Leontief production technologies, and its value here is as a bridge. A closer nonlinear input–output construction is available in Krause (1992), where technical coefficients depend on output and path stability can coexist with chaotic behaviour of individual price paths. That result is conceptually striking, but the Nishimura–Yano map makes the relation between an orthodox optimisation problem, a steady state, and chaos unusually transparent.

8.5. Equilibrium without Convergence

The bridge does not abolish equilibrium. In this model the chaotic trajectory is itself the equilibrium path: it solves the intertemporal optimisation problem, so the economy is in equilibrium in every period without ever settling. The steady state is the stationary equilibrium, one point of the same law of motion. It exists, is the unique strictly positive fixed point, and is correctly calculated; the displayed map also has the boundary fixed point $k = 0$. What fails is the additional inference that the equilibrium path must approach the stationary equilibrium. The point remains a property of the law of motion even when almost every nearby trajectory is pushed away from it.

existence $\neq$ admissibility $\neq$ local stability $\neq$ global attraction $\neq$ predictability of the path

Equilibrium analysis and chaos analysis answer different questions about the same model. The first locates invariant states and asks whether they satisfy the economic restrictions. The second follows the orbit and asks what the law of motion does away from those states. The best argument available to the Chaos Hunter is produced by the tradition he is tempted to dismiss: an intertemporal optimisation problem, solved with standard economic machinery, generates an ergodically chaotic policy function. His rebellion identifies a real omission when equilibrium is reported without dynamics; it becomes blind when it treats the omission as proof that equilibrium is meaningless.

8.6. Two Rooms, One Elephant

The deeper difficulty is institutional. Equilibrium theory and nonlinear dynamics are often practised as two competent but weakly connected worlds. Each has its journals, canonical examples, standards of proof, and characteristic objections. The equilibrium theorist can dismiss the attractor as a mathematical curiosity without empirical content; the Chaos Hunter

can dismiss the fixed point as an ideological commitment. The bibliography of this essay initially reproduced that division: it contained equilibrium sources and chaos-and-complexity sources, but no model that clearly belonged to both. It shows how easily a field of vision becomes a reading list.

Nishimura and Yano supply the missing bridge, but a bridge is not yet an empirical explanation. A phase portrait cannot show that an economy occupies the chaotic regime. The modeller must still defend the feedback, parameter range, institutions, behavioural assumptions, and observations that would distinguish deterministic chaos from noise or structural change. Dynamic traditions have developed their own standards for this task. Sterman (1987, 1988) connected simulated decision rules to experimental behaviour; Barlas (1996) and Oliva (2003) made structure, calibration, and validation part of the claim. Gallegati et al. (2006) and Ormerod (2016) applied the same demand to econophysics. Minsky (1986) and Keen (1995) show the difference between merely producing instability and locating it in balance sheets, leverage, wages, profits, and investment.

Agent-based computational economics extends the same issue from nonlinear trajectories to emergent patterns. Heterogeneous agents, learning, networks, and local interaction can generate segregation, cascades, market regularities, and institutional change (Arthur 2015; Epstein 2006; Tesfatsion 2006). The generative achievement is important, but it is the beginning of the explanation. The relevant test is not only whether the pattern can be grown. It is which rule had to bend, how far, whether the required configuration is credible, and what evidence discriminates it from rival mechanisms (Fagiolo et al. 2007). The same two rooms can reappear between equilibrium and agent-based modelling unless the bridge is made explicit.

The Chaos Hunter's rebellion is necessary because equilibrium can become a stopping rule. It becomes productive when negation gives way to reconstruction: keep the fixed point, add the law of motion, derive the fold, state the admissible parameter region, and ask how the resulting regime meets the world. The parable places six blind men around one elephant. Economics often arranges them so that they work in different rooms. Section 9 asks what changes when they return to the same table.

9. Better Tools, Wider Sight?

9.1. *The Same Blindness with a Larger Machine*

The six blind men have now touched different parts of the economics-elephant. Their local competence is real, and so is their shared temptation: a larger version of the specialised tool will supposedly reveal the whole. The promise takes many forms, including faster processors, bigger datasets, better algorithms, and more advanced mathematics. The following parallels widen the diagnosis without replacing the prank that generated it.

Lockhart's (2009) *A Mathematician's Lament* identifies the same temptation in mathematics education: greater technical difficulty can reproduce procedural fluency at a higher level without creating more understanding. Hossenfelder (2018) describes the parallel danger in theoretical physics, where elegant formal structures can deepen knowledge while making weak empirical contact harder to acknowledge. The blackboard exercise condenses both warnings into a small scene: a successful inversion encourages confidence before anyone checks the magnitude of the result or whether the coefficients describe a productive economy.

The pattern becomes clearer when each blind man receives unlimited computational resources while keeping the same unexamined starting point:

The Answer Getter receives Mathematica, a symbolic solver, or an AI assistant and obtains the output change in milliseconds. The answer can arrive with code, a graph, and a polished explanation. None of that forces the solver to ask whether the multiplier is plausible or whether the coefficient matrix satisfies the productivity condition.

The Point Equilibrium Believer evaluates a million nearby calibrations and maps the resulting sensitivity with great precision. The paper still foregrounds one benchmark point, treats the surrounding landscape as a robustness appendix, and leaves the admissibility boundary in the code.

The P-Value Voter receives one million observations. Tiny departures from the null become detectable and the asterisks multiply. The verdict also becomes more reproducible, which is a real gain. A larger sample stabilises the classification without altering what is being classified. The threshold still converts a continuous statistic into a binary one, while the substantive question remains where it stood.

The Econometrician with Perfect Data receives a terabyte of administrative records. Sampling uncertainty shrinks, while weak identification, selection, misspecified functional form, and a causal story imported from another model may survive.

The Axiomatist receives a formal verification system and proves with machine-checked rigour that aggregation preserves structure under the stated assumptions. The proof becomes stronger, while the empirical plausibility of the sameness and additivity that support it remains a separate question.

The Chaos Hunter receives a supercomputer and a decade of tick-by-tick market data. The expanded search reveals more cycles, bifurcations, and unstable regions. In his agent-based mode, the same machine supplies a larger artificial society with more agents, networks, rules, and adaptive strategies; more emergent patterns appear. Their economic meaning still depends on mechanisms, behavioural foundations, plausible parameter regions, empirical discrimination, and a bridge to the world outside the simulation.

The common pattern separates precision from understanding. Computational power can enlarge the first achievement while the second continues to depend on interpretation, comparison, and criticism. McCloskey's (1998) account of modernist methodology treats mathematical and statistical precision as the central warrant of knowledge. The classroom exercise shows how precision becomes faster blindness when specialised results never meet.

9.2. *The Exoskeleton*

Artificial intelligence strengthens the shared temptation because it can become an exoskeleton for every blind man at once. It can calculate, estimate, optimise, search specifications, generate simulations, formalise proofs, classify dynamic regimes, and narrate the result. The reduction in friction is extraordinary. The epistemic danger lies in mistaking fluency across these operations for an independent check on the assumptions that organise them.

AI is an expression of collective knowledge, or more precisely of the recorded traces of that knowledge, shaped by availability, repetition, curation, and training. This is both why it is useful and why it can hold a researcher inside a bubble. A modelling convention repeated across a large literature becomes exceptionally easy to reproduce. Messeri and Crockett (2024) warn that AI in science can foster illusions of understanding and scientific monocultures. Doshi and Hauser (2024) show that generative AI can improve individual creative output while reducing diversity across outputs. Sourati et al. (2026) review a broader homogenising effect of large language models on expression and thought.

Machine learning also contains its own methodological warning. Shortcut learning can deliver strong benchmark performance by exploiting regularities that fail when the environment changes (Geirhos et al. 2020). Underspecification allows several models to perform almost identically in the development domain and behave very differently after deployment (D'Amour et al. 2022). Internal computational success does not settle the relation between a representation and the world in which it will be used.

The location of these objections matters. They were not raised against machine learning from outside it. They were established inside the field, by its own researchers, using its own criteria. This is the third appearance of the same pattern in this essay. Sonnenschein, Mantel and Debreu proved their result inside general equilibrium theory. Hedges and Olkin established the inconsistency of vote counting inside research synthesis. Geirhos and D'Amour identify analogous limits inside machine learning. The problem is not a shortage of criticism. It is the ease with which a field can possess a criticism, teach it, and leave its routine practice unchanged.

9.3. The Collapse in the Cost of Objection

Everything said so far describes what these systems do when they are asked for an answer. They behave differently when asked for an objection, an alternative, or a comparison. The conversation recommended in this essay has always been expensive. To learn what the other five specialists would say, an economist needed five colleagues with time to look, five literatures in which he had not been trained, or a seminar that happened to contain the right objections. Skipping that step was often an honest response to a real constraint. The constraint has now largely disappeared.

For practical purposes, the marginal technical cost of constructing and exploring another model has collapsed towards zero. An assumption can be relaxed and the conclusion recomputed in a minute. A specification can be re-estimated four other ways before lunch. A small application can be built to map the region in which a model describes an admissible economy, and the researcher can play with the model rather than merely inspect its benchmark. An objection from an unfamiliar literature can be summoned in seconds and then, this part is not optional, checked against its source.

The value of this change is choice. AI does not make the elephant fully visible. It makes alternative partial views almost costless to construct, explore, and compare. The menu is real, although it is not neutral: recorded knowledge is uneven, dominant approaches are easier

to reproduce, and unfamiliar alternatives still have to be requested. Yet this is why choice matters. A default answer can preserve the bubble; a deliberately expanded menu can expose the researcher to models, mechanisms, and objections that previously lay beyond the feasible set.

Hao et al. (2026) document an adjacent version of this tension in 41.3 million natural-science papers. In their analysis, scientists engaged in AI-augmented research published 3.02 times as many papers and received 4.84 times as many citations, while the collective volume of topics studied was 4.63 per cent smaller and engagement between scientists was 22 per cent lower. These are associations within their design, not a licence to attribute every difference mechanically to AI. They nevertheless show the tension clearly: lower production costs can expand individual output while collective attention becomes narrower.

The asymmetry is simple. Fluency arrives by default; objections arrive only when someone asks. AI can become an exoskeleton for blindness or an instrument of organised doubt. Nothing in the technology decides which use prevails.

One distinction locates the scarce input. Building four alternative models is now effectively free. Knowing which of the four deserves belief is not. The bottleneck has moved from production to evaluation: internal coherence, economic credibility, comparison across model worlds, confrontation with evidence, and judgement about which differences matter.

9.4. What Conversation Adds

The parable offers its remedy through conversation. The blind men compare impressions, expose contradictions, and assemble partial truths into an account that none could reach alone. Conversation here means an organised division of epistemic labour rather than agreement.

It adds four things. First, coverage: another participant reaches a part of the animal that the first has not touched. Second, relations: a leg, tusk, and trunk cease to be isolated objects and acquire positions within a body. Third, calibration: the account reveals what it has observed, what it has not, and where its authority ends. Fourth, choice: competing accounts of the whole can be compared by asking which assumptions they require, which observations they organise, and what evidence could discriminate among them. The result is not complete knowledge. It is a more adequate and more accountable model of the whole.

The Answer Getter checks local correctness, then asks whether the number has an admissible sign, plausible magnitude, meaningful units, and an economic interpretation.

The Point Equilibrium Believer asks what object the model defines and how that object changes across plausible inputs. The P-Value Voter, recast as a statistician rather than a caricature, asks what the sample supports and with what degree of uncertainty. The Econometrician asks whether the target is identified, whether the specification matches the question, and whether causal language has been earned by the design or inherited from theory.

The Axiomatist asks which individual properties survive aggregation, which restrictions make a representative model valid, and how heterogeneity changes the result. The Chaos Hunter asks which feedbacks generate cycles, thresholds, or collapse, which initial conditions and parameter regions produce them, and what evidence connects those regions to an actual economy. In the agent-based extension, he asks which behavioural rules, interaction structures, learning processes, and institutions generate the emergent pattern, whether alternative mechanisms reproduce it, and what observations could discriminate among them. Every question is legitimate. Blindness begins when one question silences the others.

Sugden's account of credible worlds clarifies the epistemic role of comparison. Experience of an actual economy carries the authority of contact with the world, yet its mechanisms arrive entangled with many other influences. A model creates a controlled counterfactual world in which selected relationships can be followed. Inference back to an actual economy is inductive and gains force when the model world is credible as an account of what could have been true (Sugden 2000, 2009). A family of credible worlds makes that inference comparative. Varying assumptions reveals which conclusions persist, which mechanisms carry the result, and where the inference depends on one construction.

The prank presents the failure of that comparison in miniature: the blackboard accumulates many model worlds after stripping away the assumptions that distinguish them. The parameter map restores those assumptions; the Hawkins–Simon condition supplies a first test of economic credibility; evidence about coefficients, institutions, and mechanisms connects the model worlds to a target economy. Confidence emerges from a sequence rather than a single result: internal solution, economic credibility, comparison across model worlds, and confrontation with experience.

The same comparison operates between traditions and not only between calibrations. James M. Buchanan (1964) argues that economics should study exchange and the institutions through which it occurs rather than treat allocation as a solved maximisation problem. Mirowski and Nik-Khah (2017) trace a later movement towards designing market mechanisms that organise information and produce outcomes. Read together, these are not interchangeable

criticisms: they differ over whether the economist's task is to understand an open-ended process of exchange or to construct institutions that make a chosen process perform. An economist who has read only one has a description rather than a comparison. Returning to Chang's (2014) original use of the parable, competing schools are locally coherent and touch different parts of the economy.

McCloskey (1998, 2000a) provides the philosophical foundation by describing economics as a practice of argument alongside its machinery of prediction. Human argument requires a position, an objection, a revision, and a new statement. A serious policy claim should survive a sanity check, a map of uncertainty, an identification argument, a transparent bridge between theory and estimated form, comparison with credible alternatives, and an account of the process that selects or generates the observed outcome. Dialogue coordinates these tests into a disciplined synthesis.

9.5. The Conversation Inside One Head

The parable needs six men. Methodologically, however, the six perspectives need not remain in six heads. An economist at work rarely has five colleagues from other specialisms sitting beside him. The remedy cannot depend on an unusually well-staffed seminar. The six blind men can become six questions asked inside one act of inquiry.

This is the Socratic operation that matters here. A position is stated, an objection is put, the position is revised, and the revision is stated again. None of these steps requires a second person, only the willingness to become the second person oneself. The external conversation widens the available perspectives; the internal conversation creates the disposition to seek them. The researcher asks of a result he has just produced and would like to believe: What else could this be? What is this method unable to see? Which alternative would change the conclusion? What evidence would allow me to choose?

A habit of that kind is difficult to install by instruction alone. This is the point of the prank and the reason the course opens with a provocation rather than a lecture. A methodological warning delivered as a rule is recalled at an examination. The prank leaves a different trace: the memory of having finished, of the number being right, and of the room agreeing, followed by the discovery that the feeling of completion had been unearned.

The trace attaches to that feeling. The feeling recurs when a matrix inverts, a coefficient crosses a threshold, or a simulation settles down. The prank succeeds if the memory returns

at that moment, unbidden and slightly unwelcome, as a question rather than a prohibition: what is this result not showing me?

AI makes the response to that question cheap. It does not make anyone want to ask it. The entire reduction in cost is available to a researcher who never requests a counterexample and spends it instead on producing the expected conclusion faster and in better prose. The disposition to ask is the scarce input that the prank is designed to supply.

9.6. Returning to the Elephant

The blackboard began as a record of apparent arithmetic failure and ended as a map: of a parameter region crossing the boundary of economic admissibility, of information destroyed by a reporting rule, and of a stable winning category assembled from economically different objects. Nothing on it was calculated incorrectly. That is what made it worth an afternoon.

This essay is not designed to be exhausted in one reading. Its density is deliberate, but it is not a demand that students master every warning at once. The prank supplies an experiential anchor. Later courses provide the moments at which particular sections become legible: when the inverted matrix is their own, when the p-value is one they are about to report, when the representative agent erases heterogeneity, or when the stable point fails to attract. The essay is an object to return to, not a taxonomy to memorise.

What the exercise asks afterwards is small and difficult. It asks economists to leave the method in which they are competent and enter a space where they become beginners again. Tunnel vision is not a failure of intelligence. It arises when the narrow field of view is also the field in which one is trained, rewarded, and comfortable. AI has sharply reduced the technical price of stepping outside that field. Remaining inside the cocoon is increasingly a choice rather than a constraint.

A hundred observations were enough for the class to recognise something and name it confidently. They were not enough to notice that thirty of the underlying economies could not exist. Recognition is not exhaustion, and the gap between them is not closed by more observations of the same kind. Science does not require a view from nowhere or a final account of the whole. It requires partial accounts to be related, challenged, compared, and tested. Additional evidence can enlarge an account and sometimes discriminate between rival models. It cannot by itself restore relations removed by aggregation, recover a dimension the instrument excludes, or decide which model deserves belief.

The success of conversation should be stated without inflation but also without false modesty: the account gains the coverage, relations, calibration, and choice described in Section 9.4. We do not arrive at complete knowledge of the elephant. We move closer to an adequate account of it, and that is already a scientific achievement.

The recurrent danger can now be stated once more: the well-behaved statistic is not the one that answers the economic question, and its good behaviour is what makes it persuasive. Conversation interrupts that persuasion by asking what the statistic has left outside its field of vision.

That is the trace the prank is built to leave. Not the six names and not every condition of every theorem. It should leave the memory that a correct calculation once felt exactly like an answer and was not one. Years later, when an equilibrium, coefficient, p-value, representative agent, or simulation appears to close a problem, the student should hear the interruption planted in that room: What am I doing? What else could this be? What am I not seeing, and what would allow me to choose? Better fingers make each partial impression sharper. Conversation makes the impressions form a more adequate picture, without pretending that the picture is the animal itself.

References

- Ackerman, F., 2002. Still dead after all these years: Interpreting the failure of general equilibrium theory. *Journal of Economic Methodology*, 9(2), 119–139.
- American Psychological Association, 2017. Ethical principles of psychologists and code of conduct (2002, amended effective June 1, 2010, and January 1, 2017). <https://www.apa.org/ethics/code> (accessed 06.10.26)
- Amrhein, V., Greenland, S., McShane, B., 2019. Scientists rise up against statistical significance. *Nature*, 567(7748), 305–307.
- Arnsperger, C., Varoufakis, Y., 2006. What is neoclassical economics? The three axioms responsible for its theoretical oeuvre, practical irrelevance and, thus, discursive power. *Panoeconomicus*, 53(1), 5–18.
- Arthur, W.B., 2015. *Complexity and the economy*. Oxford University Press, Oxford.
- Barlas, Y., 1996. Formal aspects of model validity and validation in system dynamics. *System Dynamics Review*, 12(3), 183–210.
- Belsley, D.A., Kuh, E., Welsch, R.E., 1980. *Regression diagnostics: Identifying influential data and sources of collinearity*. Wiley, New York.
- Benjamin, D.J., Berger, J.O., Johannesson, M., Nosek, B.A., Wagenmakers, E.-J., Berk, R., et al., 2018. Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6–10.
- Buchanan, J.M., 1964. What should economists do? *Southern Economic Journal*, 30(3), 213–222.
- Caballero, R.J., 2010. Macroeconomics after the crisis: Time to deal with the pretense-of-knowledge syndrome. *Journal of Economic Perspectives*, 24(4), 85–102.
- Chang, H.-J., 2014. *Economics: The user's guide*. Pelican, London.
- Cook, R.D., 1977. Detection of influential observation in linear regression. *Technometrics*, 19(1), 15–18.
- Cumming, G., 2008. Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, 3(4), 286–300.
- D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., et al., 2022. Underspecification presents challenges for credibility in modern machine learning. *Journal of Machine Learning Research*, 23(226), 1–61.
- Debreu, G., 1970. Economies with a finite set of equilibria. *Econometrica*, 38(3), 387–392.

- Debreu, G., 1974. Excess demand functions. *Journal of Mathematical Economics*, 1(1), 15–21.
- Denis, A., 2007. The hypostatisation of the concept of equilibrium in neoclassical economics, in: Mosini, V. (Ed.), *Equilibrium in economics: Scope and limits*. Routledge, London, pp. 261–279.
- Doshi, A.R., Hauser, O.P., 2024. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28), eadn5290.
- Epstein, J.M., 2006. *Generative social science: Studies in agent-based computational modeling*. Princeton University Press, Princeton.
- Fagiolo, G., Moneta, A., Windrum, P., 2007. A critical guide to empirical validation of agent-based models in economics: Methodologies, procedures, and open problems. *Computational Economics*, 30(3), 195–226.
- Gallegati, M., Keen, S., Lux, T., Ormerod, P., 2006. Worrying trends in econophysics. *Physica A: Statistical Mechanics and Its Applications*, 370(1), 1–6.
- Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., Wichmann, F.A., 2020. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2, 665–673.
- Gelman, A., Carlin, J., 2014. Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, 9(6), 641–651.
- Gelman, A., Stern, H., 2006. The difference between ‘significant’ and ‘not significant’ is not itself statistically significant. *The American Statistician*, 60(4), 328–331.
- Gorman, W.M., 1953. Community preference fields. *Econometrica*, 21(1), 63–80.
- Greer, B., 1997. Modelling reality in mathematics classrooms: The case of word problems. *Learning and Instruction*, 7(4), 293–307.
- Hahn, F.H., 1973. *On the notion of equilibrium in economics: An inaugural lecture*. Cambridge University Press, Cambridge.
- Hao, Q., Xu, F., Li, Y., Evans, J., 2026. Artificial intelligence tools expand scientists’ impact but contract science’s focus. *Nature*, 649, 1237–1243.
- Hargreaves Heap, S.P., Varoufakis, Y., 2004. *Game theory: A critical text* (2nd ed.). Routledge, London.
- Hawkins, D., Simon, H.A., 1949. Note: Some conditions of macroeconomic stability. *Econometrica*, 17(3), 245–248.
- Hedges, L.V., Olkin, I., 1980. Vote-counting methods in research synthesis. *Psychological Bulletin*, 88(2), 359–369.

- Hicks, J.R., 1939. Value and capital: An inquiry into some fundamental principles of economic theory. Clarendon Press, Oxford.
- Hossenfelder, S., 2018. Lost in math: How beauty leads physics astray. Basic Books, New York.
- Ioannidis, J.P.A., Stanley, T.D., Doucouliagos, H., 2017. The power of bias in economics research. *The Economic Journal*, 127(605), F236–F265.
- Jorgenson, D.W., 1960. A dual stability theorem. *Econometrica*, 28(4), 892–899.
- Jorgenson, D.W., 1961. Stability of a dynamic input–output system. *The Review of Economic Studies* 28 (2), 105–116.
- Keen, S., 1995. Finance and economic breakdown: Modeling Minsky’s ‘financial instability hypothesis’. *Journal of Post Keynesian Economics*, 17(4), 607–635.
- Kirkland, P.K., McNeil, N.M., 2021. Question design affects students’ sense-making on mathematics word problems. *Cognitive Science*, 45(4), e12960.
- Kirman, A.P., 1989. The intrinsic limits of modern economic theory: The emperor has no clothes. *The Economic Journal*, 99(395), 126–139.
- Kirman, A.P., 1992. Whom or what does the representative individual represent? *Journal of Economic Perspectives*, 6(2), 117–136.
- Krause, U., 1992. Path stability of prices in a nonlinear Leontief model. *Annals of Operations Research*, 37(1), 141–148.
- Lakens, D., Adolphi, F.G., Albers, C.J., Anvari, F., Apps, M.A.J., Argamon, S.E., et al., 2018. Justify your alpha. *Nature Human Behaviour*, 2(3), 168–171.
- Leamer, E.E., 1983. Let’s take the con out of econometrics. *American Economic Review*, 73(1), 31–43.
- Leibenstein, H., 1976. Beyond economic man: A new foundation for microeconomics. Harvard University Press, Cambridge, MA.
- Leontief, W., 1986. Input-output economics (2nd ed.). Oxford University Press, Oxford.
- Lockhart, P., 2009. A mathematician’s lament: How school cheats us out of our most fascinating and imaginative art form. Bellevue Literary Press, New York.
- Lorenz, E.N., 1963. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2), 130–141.
- Manski, C.F., 2011. Policy analysis with incredible certitude. *The Economic Journal*, 121(554), F261–F289.

- Mantel, R.R., 1974. On the characterization of aggregate excess demand. *Journal of Economic Theory*, 7(3), 348–353.
- Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. *Microeconomic theory*. Oxford University Press, Oxford.
- May, R.M., 1976. Simple mathematical models with very complicated dynamics. *Nature*, 261, 459–467.
- McCloskey, D.N., 1998. *The rhetoric of economics* (2nd ed.). University of Wisconsin Press, Madison.
- McCloskey, D.N., 2000a. *How to be human: Though an economist*. University of Michigan Press, Ann Arbor.
- McCloskey, D.N., 2000b. Other things equal: How to be scientific in economics. *Eastern Economic Journal*, 26(2), 241–246.
- Messeri, L., Crockett, M.J., 2024. Artificial intelligence and illusions of understanding in scientific research. *Nature*, 627, 49–58.
- Minsky, H.P., 1986. *Stabilizing an unstable economy*. Yale University Press, New Haven.
- Mirowski, P., 1989. *More heat than light: Economics as social physics, physics as nature's economics*. Cambridge University Press, Cambridge.
- Mirowski, P., 1991. The when, the how and the why of mathematical expression in the history of economic analysis. *Journal of Economic Perspectives*, 5(1), 145–157.
- Mirowski, P., Nik-Khah, E., 2017. *The knowledge we have lost in information: The history of information in modern economics*. Oxford University Press, Oxford.
- Nishimura, K., Stachurski, J., 2004. Discrete time models in economic theory. *CUBO, A Mathematical Journal*, 6(1), 187–207.
- Nishimura, K., Yano, M., 1995. Nonlinear dynamics and chaos in optimal growth: An example. *Econometrica*, 63(4), 981–1001.
- Oliva, R., 2003. Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research*, 151(3), 552–568.
- Ormerod, P., 2016. Ten years after ‘Worrying trends in econophysics’: Developments and current challenges. *The European Physical Journal Special Topics*, 225(17), 3281–3291.
- Ortmann, A., Hertwig, R., 2002. The costs of deception: Evidence from psychology. *Experimental Economics*, 5(2), 111–131.

- Palm, T., 2008. Impact of authenticity on sense making in word problem solving. *Educational Studies in Mathematics*, 67(1), 37–58.
- Plato, 1925. *Gorgias*. Translated by W.R.M. Lamb. Harvard University Press, Cambridge, MA (original work ca. 380 BCE).
- Reusser, K., Stebler, R., 1997. Every word problem has a solution: The social rationality of mathematical modeling in schools. *Learning and Instruction*, 7(4), 309–327.
- Rizvi, S.A.T., 2006. The Sonnenschein-Mantel-Debreu results after thirty years. *History of Political Economy*, 38(Suppl. 1), 228–245.
- Samuelson, P.A., 1947. *Foundations of economic analysis*. Harvard University Press, Cambridge, MA.
- Schoenfeld, A.H., 1988. When good teaching leads to bad results: The disasters of ‘well-taught’ mathematics courses. *Educational Psychologist*, 23(2), 145–166.
- Skemp, R.R., 1976. Relational understanding and instrumental understanding. *Mathematics Teaching*, 77, 20–26.
- Sonnenschein, H., 1972. Market excess demand functions. *Econometrica*, 40(3), 549–563.
- Sourati, Z., Ziabari, A.S., Dehghani, M., 2026. The homogenizing effect of large language models on human expression and thought. *Trends in Cognitive Sciences*, advance online publication, <https://doi.org/10.1016/j.tics.2026.01.003>.
- Sterman, J.D., 1987. Testing behavioral simulation models by direct experiment. *Management Science*, 33(12), 1572–1592.
- Sterman, J.D., 1988. Deterministic chaos in models of human behavior: Methodological issues and experimental results. *System Dynamics Review*, 4(1–2), 148–178.
- Sugden, R., 2000. Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, 7(1), 1–31.
- Sugden, R., 2009. Credible worlds, capacities and mechanisms. *Erkenntnis*, 70(1), 3–27.
- Tesfatsion, L., 2006. Agent-based computational economics: A constructive approach to economic theory, in: Tesfatsion, L., Judd, K.L. (Eds.), *Handbook of computational economics: Vol. 2. Agent-based computational economics*. Elsevier/North-Holland, Amsterdam, pp. 831–880.
- Varoufakis, Y., 2014. *Economic indeterminacy: A personal encounter with the economists’ peculiar nemesis*. Routledge, London.

- Verschaffel, L., De Corte, E., Borghart, I., 1997. Pre-service teachers' conceptions and beliefs about the role of real-world knowledge in mathematical modelling of school word problems. *Learning and Instruction*, 7(4), 339–359.
- Verschaffel, L., De Corte, E., Lasure, S., 1994. Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction*, 4(4), 273–294.
- Verschaffel, L., Greer, B., De Corte, E., 2000. Making sense of word problems. Swets & Zeitlinger, Lisse.
- Wasserstein, R.L., Lazar, N.A., 2016. The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133.

Appendix A. The Threshold, the Tail, and the Finite Grid

This appendix records the calculations behind Section 7. It derives the heterogeneity threshold at which the continuous coefficient population first reaches $D = 0$, states the heavy-tail result, and distinguishes that population experiment from the finite classroom grid.

A.1. The Threshold and the Tail

For e_1 and e_2 measured as deviations from the baseline calibration, the two-sector determinant and the first output solution are

$$D(e_1, e_2) = 0.006 - 0.54e_1 - 0.40e_2, \quad \Delta x_1(e_1, e_2) = \frac{0.60 - 40e_2}{D(e_1, e_2)}$$

Let (e_1, e_2) be uniformly distributed on $[-h, h]^2$. The minimum determinant on this square is $0.006 - 0.94h$, so the support first reaches zero at $h^* = \frac{0.006}{0.94} \approx 0.00638$. For $h < h^*$, D remains bounded away from zero and the moments of Δx_1 are finite. For $h > h^*$, the density of D is positive in a neighbourhood of zero. Except at isolated numerator cancellations, division by D then produces tails with density proportional to $|x|^{-2}$. Consequently $E|\Delta x_1|$ is infinite and the ordinary mean $E(\Delta x_1)$ is undefined.

At $h = 0.004$, numerical integration gives a representative solution of 100.00, a median of 100.00, and a mean of approximately 105.14; all draws satisfy Hawkins–Simon. At $h = 0.05$, the representative solution remains 100.00, the population median is approximately 37.66, and the non-productive share is approximately 0.3889. The population mean has no finite value.

A.2. The Finite Classroom Grid and the Continuous Population

The classroom contains one hundred discrete assignments. None has $D = 0$: the smallest absolute determinant is 0.0004. Every individual algebraic solution, and hence the mean and median of any non-empty finite sample of those solutions, is finite. Thirty assigned matrices nevertheless fail the Hawkins–Simon condition, and observations near the boundary dominate magnitude-based summaries.

This finiteness does not guarantee that a matrix constructed from average coefficients is invertible. For example, the assigned pairs $(e_1, e_2) = (-0.04, -0.01)$ and $(0.04, 0.04)$ have determinants 0.0316 and -0.0316 , respectively, so each individual system is invertible. Their mean parameter pair is $(0, 0.015)$, for which $D = 0$. The representative system is singular and has no unique solution, even though both individual solutions and their arithmetic mean are finite.

Averaging solved models and solving a model built from average coefficients therefore remain distinct operations even on a finite grid.

The continuous experiment fills the parameter square rather than selecting one hundred points from it. Once h exceeds h^* , the support includes $D = 0$ and arbitrarily close neighbourhoods on both sides. This is the setting in which the population-moment statement applies. The discrete grid demonstrates sensitivity, compression, and economic inadmissibility; the continuous population establishes the non-existence of the mean. Keeping the two designs separate preserves both claims.



UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
44/50 DŁUGA ST.
00-241 WARSAW
WWW.WNE.UW.EDU.PL
ISSN 2957-0506