

Development Economics after the Experimental Turn: A Critical Review of a Displaced Object of Inquiry**Philippe Patrick ANDRIANJAFY*****University of Toliara and University of Ambovombe Androy- Madagascar****aphilippepatrcik@gmail.com****<https://orcid.org/0009-0005-9813-2849>****<https://doi.org/10.63939/JSS.2026-Vol10.N41.165-195>****Received: 11/08/2026, Accepted: 14/09/2026, Published: 18/09/2026**

Abstract: The experimental turn has been argued over as a quarrel about method. This article makes a different claim: what the turn produced was a substantial reweighting of the object of inquiry in development economics, programme-level causal effects becoming the dominant form in which development problems were rendered researchable. The review draws on some sixty texts published between 1995 and 2025, selected on three criteria: the founding and handbook statements of the experimental programme, the peer-reviewed methodological criticism it drew in general and in development journals, and the post-2015 work on aggregation, general equilibrium and scaling. It works from an explicit criterion of demarcation: an object has been displaced when the questions the apparatus cannot handle become markedly less central in the field's journals, not when they receive different answers there. Four shifts are documented. The unit of analysis moves from the trajectory of an economy to the average effect of a programme. The state moves from central object of explanation to implementing institution. The time horizon moves from structural transformation to the two or three years of a survey cycle. Causal credibility comes to rest more on assignment than on an explicit account of mechanisms. The magnitudes give the measure of the shift: the J-PAL network reports a cumulative portfolio rising from 103 randomised evaluations in 2003 to 2,566 in 2025 across 124 countries (J-PAL, 2026), and by 2015 empirical work, not experimental work specifically, accounted for more than 90 per cent of weighted publication pages in development. From this the article draws a proposition about the past decade: the corrections devised since 2015, Bayesian aggregation, general equilibrium estimation, the political economy of scaling, put back one by one the elements the apparatus had set aside, namely structure, aggregation and the state. Two results measure that return: the local transfer multiplier of 2.5 estimated on cash grants in Kenya, and the undetectable effect of contract teachers hired by the Kenyan administration against 0.18 standard deviations for the same contract delivered by an NGO. A bibliometric probe on 18,592 articles published in seven development journals between 1990 and 2025 puts figures on the reweighting: the share carrying an experimental vocabulary rises from 0.71 to 9.63 per cent of annual output between the 1990s and 2016-2025, reaching 18.4 per cent in the Journal of Development Economics, while the structural vocabulary follows a U-shaped path, halving during the 2000s and recovering afterwards everywhere except in that same journal. The epistemological consequence falls on the global South: the evidence available to policymakers there is composed of programme-level effects measured in the few countries cheapest to survey, while the questions on which their macroeconomic and industrial choices are decided remain, by construction of the method, the least researched.

Keywords: development economics; randomised controlled trials; external validity; economic methodology; structural transformation; political economy of evaluation.

JEL codes: B41, O10, O12, C93, B23

© 2026, [Philippe Patrick ANDRIANJAFY], licensee Democratic Arab Center. This article is published under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which permits non-commercial use, distribution, and reproduction in any medium or format, provided the original work is properly cited. Official License Link: <https://creativecommons.org>

*Corresponding author

اقتصاد التنمية بعد المنعطف التجريبي: مراجعة نقدية لموضوع بحث مزاح

فيليب باتريك أندريانجافي *

جامعة توليوار وجامعة أميوفو أمبوفومبي أندروي - مدغشقر

aphilippepatrcik@gmail.com / <https://orcid.org/0009-0005-9813-2849><https://doi.org/10.63939/JSS.2026-Vol10.N41.165-195>

Received: 11/08/2026, Accepted: 14/09/2026, Published: 18/09/2026

ملخص: لقد حدث الجدل حول المنعطف التجريبي باعتباره نزاعاً حول المنهج. إلا أن هذا المقال يطرح ادعاءً مختلفاً: وهو أن ما أنتج هذا التحول كان إعادة وزن جوهرية لموضوع البحث في اقتصاد التنمية، حيث أصبحت التأثيرات السببية على مستوى البرامج هي الشكل المهيمن الذي صيغت به مشكلات التنمية لتصبح قابلة للبحث. تعتمد هذه المراجعة على نحو ستين نصاً نُشرت بين عامي 1995 و2025، تم اختيارها بناءً على ثلاثة معايير: البيانات التأسيسية والبيانات المرجعية للبرنامج التجريبي، والنقد المنهجي المحكم الذي تعرض له عموماً وفي المجالات التنموية، والأعمال المنشورة بعد عام 2015 بشأن التجميع، والتوازن العام، وتوسيع النطاق. سوف ننطلق من معيار تمييز صريح: يُزاح موضوع البحث عندما تصبح الأسئلة التي تعجز الأدوات المنهجية عن معالجتها أقل مركزية بشكل ملحوظ في مجلات هذا التخصص، وليس عندما تتلقى إجابات مختلفة فيها. وقد تم توثيق أربعة تحولات: تحركت وحدة التحليل من مسار الاقتصاد إلى متوسط تأثير البرنامج. وتحركت الدولة من كونها موضوع التفسير المركزي إلى مؤسسة منفذة. وتحرك الأفق الزمني من التحول الهيكلي إلى السنتين أو الثلاث سنوات الخاصة بدورة المسح. وأصبح الاعتماد في المصادقية السببية يركز على التوزيع العشوائي للمجموعات أكثر من ارتكازه على التفسير الصريح للأليات. وتعطي الأرقام مقياساً لهذا التحول: إذ يشير معهد عبد اللطيف جميل لمكافحة الفقر (J-PAL) إلى ارتفاع محفظته التراكمية من 103 تقييمات عشوائية في عام 2003 إلى 2,566 تقييم في عام 2025 شملت 124 دولة (J-PAL, 2026). وبحلول عام 2015، أصبحت البحوث التطبيقية (وليس العمل التجريبي على وجه التحديد) تستحوذ على أكثر من 90% من صفحات النشر المرجحة في مجال التنمية. ومن هذا المنطلق، يطرح المقال فرضية حول العقد الماضي: وهي أن التصحيحات التي تم ابتكارها منذ عام 2015 (مثل التجميع البايزي، وتقدير التوازن العام، والاقتصاد السياسي لتوسيع النطاق) أعادت واحداً تلو الآخر العناصر التي استبعدتها الأدوات المنهجية سلفاً، ونعني بها البنية، والتجميع، والدولة. وهناك نتيجتان تقيسان هذه العودة: مضاعف التحويل المحلي البالغ 2.5 الذي تم تقديره على المنح النقدية في كينيا، والأثر غير الملحوظ للمعاقدن من المعلمين الذين وظفتهم الإدارة الكينية مقابل 0.18 انحراف معياري لنفس العقد عندما قدمته منظمة غير حكومية. ويضع التقصي الببليومتري لـ 18,592 مقالاً نُشرت في سبع مجلات تنموية بين عامي 1990 و2025 أرقاماً دقيقة حول إعادة الوزن هذه: إذ ارتفعت حصة المقالات التي تحمل مصطلحات تجريبية من 0.71% إلى 9.63% من الإنتاج السنوي بين التسعينيات والفترة (2016-2025)، لتصل إلى 18.4% في مجلة اقتصاد التنمية (Journal of Development Economics)، في حين اتبع المعجم البنيوي مساراً على شكل حرف U، حيث انخفض إلى النصف خلال العقد الأول من القرن الحادي والعشرين ثم تعافى بعد ذلك في كل مكان باستثناء تلك المجلة نفسها. أما النتيجة الإستمولوجية (المعرفية) فتقع على عاتق الجنوب العالمي: إذ إن الأدلة المتاحة لصناع السياسات هناك باتت تتكون من تأثيرات على مستوى البرامج تم قياسها في البلدان القليلة الأرخص تكلفة في إجراء المسوح، في حين تظل الأسئلة التي تتحدد بناءً عليها خياراتهم الاقتصادية الكلية والصناعية هي الأقل بحثاً، بموجب طبيعة وبنية المنهج نفسه.

الكلمات المفتاحية: اقتصاد التنمية، التجارب العشوائية المحكومة، الصلاحية الخارجية، المنهجية الاقتصادية، التحول الهيكلي، الاقتصاد السياسي للتقييم.

1. Introduction

When the Sveriges Riksbank Prize went to Abhijit Banerjee, Esther Duflo and Michael Kremer in 2019, it honoured a way of doing development economics that twenty years earlier did not exist as a recognisable research programme. The three Nobel lectures (Duflo, 2020; Banerjee, 2020; Kremer, 2020) advance a claim that is convergent and, in its wording, modest: field experimentation is not a theory of development, it is a way of finding out whether an intervention does what it is said to do. That modesty disarmed a good part of the opposition. It also obscured what was at stake.

The controversy opened by Heckman and Smith (1995) and carried further by Deaton (2010), Barrett and Carter (2010), Acemoglu (2010), Ravallion (2020) and Bédécarrats, Guérin and Roubaud (2019) was conducted almost entirely on methodological ground: internal against external validity, local average effect against structural parameter, hierarchy of evidence against methodological pluralism. That framing produced a large literature and a settled position, which can be put in a sentence. A randomised trial identifies an effect cleanly in a given site, and by itself says nothing beyond that. Most contemporary contributors accept that a single trial does not by itself establish the transportability of its result beyond the setting studied. The methodological controversy has stabilised around a set of recurring distinctions, without the substantive disagreements having been settled.

What is not closed lies elsewhere. A method does not merely answer a field's questions differently; it selects which questions can be asked at all. The experimental apparatus requires a unit of assignment, a credible control group, an observation window compatible with a research grant, and an intervention delimited enough to be administered to one part of a population and withheld from another. These four requirements are not neutral technicalities. They are a filter. What passes through the filter occupies the centre of the field; what is held back loses its centrality in the journals, then in the dissertations, then in the syllabuses.

My thesis is this. Between the late 1990s and the mid-2010s, development economics reweighted its object, and the reweighting was treated as a change of method. By displacement of the object of inquiry I mean a change in the questions a field treats as central, which is distinct from a

change in the techniques used to answer them, and which does not require that any question disappear. The historical object of the field, the explanation of durable gaps in living standards between economies and of the conditions under which productive structures are transformed, was not collectively refuted as a research agenda, even though particular models and claims were successfully challenged. It was marginalised, reconfigured, and in part relocated to adjacent subfields, because the dominant apparatus does not take it. The proposition is interpretive in its general form, and Section 9 submits the part of it that can be counted to a measurement of my own on 18,592 articles.

Three limits bound the argument. It is not a brief against experimentation: the gains in identification are real and cumulative, and on a well-defined set of questions the discipline is the better for them. Nor is the claim that experimenters are unaware of the objections raised against them; some of the most lucid pages on external validity and on scaling are written by the experimenters themselves (Banerjee et al., 2017; Muralidharan and Niehaus, 2017; Duflo, 2017). Finally, the recent period, roughly since 2015, is treated here not as confirmation of the diagnosis but as its complication: the successive corrections reinstate the displaced object, in a form the apparatus constrains.

The article is organised as follows. Section 2 sets out how the corpus was assembled and states the criterion of demarcation. Section 3 reconstructs the genealogy of the displacement. Section 4 records what the apparatus has actually delivered. Section 5 sorts the limits established in the literature and offers a status report on the objections. Section 6 states the displacement and offers a typology along four dimensions. Section 7 examines the recompositions now under way. Section 8 discusses the hierarchy of questions and what it implies for research conducted on African economies. Section 9 puts the first condition of the criterion to a bibliometric test. Section 10 concludes and states the limits of the exercise.

Three levels not to be conflated. The argument involves three distinct series, which the reader should keep apart. The first comprises the four technical requirements of the experimental apparatus: an assignable unit, a credible control group, an observation window compatible with a research grant, a delimited intervention. The second comprises the three analytical conditions to which the displacement thesis is submitted, stated

in Section 2 and taken up in Section 6. The third comprises the four dimensions along which the displacement is ordered, set out in Section 6.4: unit of analysis, agent, time horizon, conception of causality. The requirements explain the displacement; the conditions serve to establish it; the dimensions describe it.

2. Method, corpus and criterion of demarcation

This review is organised around a thesis and makes no claim to the exhaustiveness of a systematic review of the PRISMA type. The corpus, some sixty references, was assembled along three lines. The first gathers the founding texts of the experimental programme and its canonical expositions: Duflo, Glennerster and Kremer (2007), Banerjee and Duflo (2009), the three Nobel lectures of 2020. The second collects the critical literature published in general and in development journals, from Heckman and Smith (1995) through to Ravallion (2020), including the French-language contribution (Labrousse, 2010, 2016; Jatteau, 2013; Bédécarrats, Guérin and Roubaud, 2019). The third brings together the post-2015 work on aggregation, general equilibrium and scaling, which constitutes the current state of the programme.

Two inclusion criteria were applied. A text had to have appeared in a peer-reviewed journal, a reference handbook, or with an academic publisher. It had to bear on development economics, which excludes the general methodological literature on causal inference except where that literature is explicitly taken up within the field (Imbens, 2010). Unpublished working papers were left aside, with two signposted exceptions where the text carries authority in the controversy.

The delicate point is the criterion of demarcation. To assert that a field has changed its object requires more than a count of frequencies: the rise of a method does not by itself establish the disappearance of an object. The criterion adopted here has three parts. A displacement is established when questions less compatible with the apparatus become markedly less central in the field's journals, and not merely receive different answers there; when the selection of sites and interventions follows the availability of an assignment mechanism rather than the magnitude of what is at stake; and when results of the apparatus are read as statements about development although they are statements about programmes. Section 6 takes the three in turn.

This criterion carries a methodological cost that has to be owned. It obliges one to document absences, which is a more fragile exercise than the recording of presences. The indicators used are therefore of two kinds. The first are published counts, bearing on volumes of evaluations and on the composition of journals (Cameron, Mishra and Brown, 2016; Angrist et al., 2017). The second are results internal to the experimental programme that reveal the structure of the selection, foremost among them Vivalt's (2020) finding that programmes implemented by government agencies yield systematically smaller effects than the same programmes delivered by NGOs or by academic teams. That regularity holds across studies whose implementers were not randomly assigned: it is a robust association, not a causal identification.

3. Genealogy: three objects in succession

3.1 The initial object: accumulation, structure, coordination

Development economics took shape after 1945 around a problem of intertemporal allocation and coordination. Rosenstein-Rodan (1943) formulated the complementarity of investments and the pecuniary externality that makes an isolated investment unprofitable where simultaneous investment would pay. Lewis (1954) gave productive dualism its canonical statement and made the reallocation of labour from subsistence to the capitalist sector the engine of accumulation. The object was clear enough: the productive structure of an economy, its transformation, and the institutional and financial conditions of that transformation.

Two properties of this object deserve notice, because both will vanish. The unit of analysis is the national economy, not the individual or the household. And the causal mechanism is explicitly one of general equilibrium: what makes an investment profitable is what other investors do; what makes migration profitable is the productivity differential between sectors.

The neoclassical critique of the 1970s and 1980s attacked the programme at its weakest joint, its confidence in the capacity of the state to carry out the coordination. Sen (1983) registered the retreat while defending the pertinence of the original question. Krugman (1992) offered a retrospective reading: the programme was not refuted, it was abandoned for want of a formalisation acceptable by the standards of the discipline, a

diagnosis that belongs as much to the sociology of the field as to epistemology. The observation is worth keeping in view. This is not the first time development economics has changed its object for reasons that have to do with its instruments rather than with its results.

3.2 The interregnum of cross-country regressions

The 1990s substituted for the theory of structural transformation an econometric programme built on international comparison. Barro (1991) fixed the format: growth regressed on a set of initial conditions over a panel of countries. The programme produced a mass of results and, before long, a demonstration of their fragility. Rodríguez and Rodrik (2000) showed, on trade openness, that the most cited findings do not survive scrutiny of the measures and the specifications. Durlauf, Johnson and Temple (2005) drew up the technical balance sheet: candidate regressors outnumbering observations, endogeneity throughout, parameter heterogeneity across countries.

Acemoglu, Johnson and Robinson (2001) proposed a way out through historical instrumentation, with the success one knows and the objections Deaton (2010) would raise about instrument validity and about what such instruments identify. The point that matters here lies elsewhere. By the end of the 1990s the field held an ambitious object, the deep determinants of development, and an inferential apparatus that the discipline itself had shown incapable of carrying it. The experimental turn settled into that gap. It did not conquer occupied territory; it moved into territory that had been vacated.

3.3 The apparatus takes hold

The sequence was short. Kremer and his coauthors ran the first randomised evaluations of educational and health interventions in Kenya, of which Miguel and Kremer (2004) remains the canonical instance, less for its findings on deworming than for its demonstration that treatment externalities can be estimated by randomising the level of assignment. Duflo (2001) showed, on school construction in Indonesia, that returns to education can be identified from an exogenous variation in public policy. Duflo, Glennerster and Kremer (2007) wrote the manual of procedure in the Handbook of Development Economics; Banerjee and Duflo (2009) supplied its epistemological justification in the Annual Review of Economics.

The magnitudes of the shift are on record. Cameron, Mishra and Brown (2016) count 2,259 impact evaluations published between 1981 and 2012, with a marked acceleration after 2008. The J-PAL network, which does not exhaust the field but forms its institutional core, reports a portfolio rising from 103 randomised evaluations in 2003 to 2,566 in 2025, carried out in 124 countries by more than 390 affiliated professors (J-PAL, 2026). These counts come from the network's institutional records rather than from a bibliographic census; they refer to a cumulative stock whose definition may have varied over the years, and they should be read as indicators of institutional growth rather than as a homogeneous time series. Angrist et al. (2017) measure the transformation from the side of the journals: the empirical share of publications, weighted by the standing of the outlet, rises from about one third in the early 1980s to more than 55 per cent by 2015 for the discipline as a whole, and exceeds 90 per cent in labour and development, the two fields where the swing is most complete. Their measure concerns whether work is empirical, not whether it is experimental: it establishes the empiricisation of the field, a necessary condition for the shift described here, and not that shift itself.

The rate of adoption is that of a technology, not that of a theory. The remark belongs to the sociology of science, but it has analytical force: technologies spread according to their cost of use and their yield in academic recognition, not according to their fit with an object.

4. What the apparatus has delivered

The positive record has to be stated precisely, or the critique collapses into a matter of principle.

Identification. On the question it treats, the randomised trial solves a problem that thirty years of applied econometrics had not solved. The comparison between treated and untreated rests on an assignment whose law the researcher knows. Imbens (2010), answering Deaton, put the defensible version of the argument: a cleanly estimated local effect is worth more than a structural parameter identified through untestable restrictions. The force of the argument lies not in any intrinsic superiority of randomisation but in the transparency of what it requires.

The discipline of evidence. The programme imported into applied economics a set of practices whose reach extends well beyond experimentation: pre-registration of hypotheses, analysis plans lodged

before collection, publication of data and code, trial registries. Christensen and Miguel (2018) assess both the transformation and its limits. The asymmetry is worth noting: these practices benefit work that has nothing experimental about it.

The destruction of settled beliefs. Several results have closed debates that had run for decades. The series of seven randomised evaluations of microcredit, aggregated by Meager (2019) in a Bayesian hierarchical model, establishes that the average effect on household consumption and business activity is unlikely to be transformative and may be negligible, and that true heterogeneity across sites is moderate, roughly 60 per cent of observed heterogeneity being sampling variation. For development policy the implication is considerable: it challenges the claim that standard microcredit is, on average, a transformative anti-poverty instrument, a claim on which a substantial share of anti-poverty finance had rested for twenty years. It does not disqualify credit as an instrument, a question that Section 6.3 shows these trials do not address.

The link to decision. The graduation programme, evaluated simultaneously in six countries (Banerjee et al., 2015), shows that a coordinated package of asset transfers, coaching and access to savings produces gains that persist a year after support ends. Here the apparatus supplies what no other instrument supplied: an estimated effect, a cost per beneficiary, and therefore a ratio comparable across interventions.

These gains are solid. They share one property, which is where the difficulty begins: all of them bear on programmes, that is, on objects created in order to be evaluated.

5. The established limits, and their standing

The critical literature has produced a set of objections, and it is worth sorting them by how far each has been dealt with rather than into the absorbed and the intact.

5.1 External validity is not a problem of method but a problem of theory

The classical objection, put by Deaton (2010) and systematised by Deaton and Cartwright (2018), is not that a result obtained in Kenya fails to hold in Bangladesh. It is that the statement "treatment T produces effect E" is not yet a scientific proposition until one has specified the causal structure that makes E possible, and about that structure randomisation says nothing. Cartwright and Hardie (2012) give the argument its operational

form: to transport a result one must know the auxiliary factors whose presence is required, and knowledge of those factors never comes from the trial itself.

The sharpest empirical support is Vivalt's (2020). Across the impact evaluations assembled in AidGrade she establishes substantial heterogeneity in effects and, more tellingly, a result the discipline has not sufficiently pondered: programmes implemented by government agencies produce smaller effects than the same programmes implemented by NGOs or by research teams, a gap that survives controls for sample size. Since implementers are not randomly assigned across studies, this comparison identifies an association rather than a causal effect of the operator's identity, which weakens the inference without weakening the regularity. Bold et al. (2018) provide the most striking illustration. In Kenya, a fixed-term teaching contract offered by an international NGO raised pupils' test scores by 0.18 standard deviations, while the contract that was identical on paper, offered by the Kenyan government, produced an effect indistinguishable from zero. Observable teacher characteristics explain little of the gap. The authors attribute it to bureaucratic and political opposition, to implementation delays, and to a differential reading of contract terms that were on paper the same, an interpretation their design renders plausible without identifying each of these mechanisms separately.

A second mechanism compounds the difficulty. Allcott (2015) shows, on an energy programme rolled out across a large number of American sites, that the sites willing to host an evaluation are not drawn at random from the population of possible sites, and that the effects estimated on the early sites overstate what is obtained later elsewhere. The bias does not concern assignment within the site, which remains random; it concerns the selection of the site itself, a dimension the protocol does not control.

This shifts the nature of the problem. The question is no longer whether an effect travels from one context to another. It is that the object evaluated, a programme delivered by a research operator, is not the object public policy needs, a programme delivered by an administration. The two differ not by context but by agent.

5.2 Scaling up

Muralidharan and Niehaus (2017) and Banerjee et al. (2017) treated this question from inside the programme, with a lucidity that makes their

papers more useful than most external criticism. They identify the channels through which an effect estimated on a pilot sample ceases to hold at scale: constraints on factor supply, movements in equilibrium prices, dilution of implementation quality, political reaction. The practical conclusion they draw, that one should experiment directly at scale, is coherent. It comes at a price: it brings the unit of analysis back to the administrative unit, which is the level at which general equilibrium mechanisms and political economy become decisive again.

5.3 General equilibrium

Acemoglu (2010) had stated the theoretical objection: a partial effect identified on a sample tells us nothing about the effect of the same policy applied to the whole population once prices, wages or norms adjust. The objection was taken seriously and answered experimentally. Egger et al. (2022) distributed about USD 1,000 to more than 10,500 households across 653 randomised Kenyan villages, an implied fiscal shock of more than 15 per cent of local GDP, and estimate a local transfer multiplier of 2.5, with substantial gains to non-recipient households and little local inflation.

The work is remarkable, and what it implies should be measured. To answer a question of general equilibrium it was necessary to randomise entire markets, to command an unusually large budget, and to accept that the unit of analysis would no longer be the household. General equilibrium interactions are reinstated at the level of markets and villages, at the cost of a design that the great majority of research teams cannot reproduce. The estimate remains a local general equilibrium magnitude and should not be equated with a national macroeconomic one. On this question the experimental programme resolves the difficulty by becoming a programme of general equilibrium analysis at the scale of local markets.

5.4 Statistical fragility

Young (2019) re-examines 53 experimental papers published in the journals of the American Economic Association using randomisation inference. In the average paper, such tests find 13 to 22 per cent fewer significant results than the authors' own methods for individual treatment effects, and 33 to 49 per cent fewer for joint tests of several effects reported in the same table. Ioannidis, Stanley and Doucouliagos (2017)

document, for their part, the insufficient statistical power of a large share of empirical economics and the selection bias that follows from it.

Neither paper targets experimentation as such; both reach the statistical practice of the discipline. Their bearing here is precise. The hierarchy of evidence that places the randomised trial at the top presupposes that experimental results are robust, and the size of the reductions Young measures forbids treating that presupposition as settled.

5.5 Data and execution

The replication carried out by Bédécarrats, Guérin, Morvant-Roux and Roubaud (2019) on the Moroccan trial of Crépon, Devoto, Duflo and Parienté (2015) addresses ground the methodological controversy rarely covers: the quality of the data the protocol itself produces. Low take-up, contamination across arms, internal inconsistencies in the files. Their conclusion, developed in Bédécarrats, Guérin and Roubaud (2019) and in the 2020 collective volume, is that internal validity is not a property guaranteed by the assignment plan but a property of execution, and that execution is seldom documented at the level of detail that would allow a judgement.

The French-language contribution was the first to tie this critique of execution to a political economy of the apparatus. Labrousse (2010) read it as an instrument of proof and of government; Jatteau (2013) documented ethnographically the conditions under which trials are produced in the field; Labrousse (2016) compared the approach with Ostrom's to show what technique alone cannot establish. This line of work was long treated as peripheral by the anglophone literature. The volume published by Oxford University Press in 2020, where Ravallion (2020) appears alongside the French authors, marks its entry into the central discussion.

5.6 Taking stock of the controversy

Twenty-five years of argument have produced a sorting that is worth making explicit, since repeating every objection indiscriminately weakens those that hold.

The sorting does not run between refuted and intact objections, which would be too convenient, but by degree of treatment. Figure 1 gives the state of play.

Objection	Status
Absence of explicit theory	Partially absorbed
Scaling up	Acknowledged, unresolved
General equilibrium effects	Partially addressed
External validity and transportability	Open
Role and capacity of the state	Insufficiently addressed
Ethics and power relations	Institutionalised, unresolved
Selection of questions by the apparatus	Barely addressed

Figure 1. Status of the main objections addressed to the experimental programme

Source: author's elaboration.

Three qualifications accompany that reading. The charge of atheoreticism is only partly absorbed: the programme has taken on the specification of mechanisms and the testing of behavioural predictions, but Card, DellaVigna and Malmendier (2011) remind us that the proportion of experiments designed to discriminate between theories remains low, and the theoretical question is not thereby closed. The existence of trials conducted across entire administrative jurisdictions does not refute the objection about scale: it shows the exercise to be feasible under conditions and with considerable means, which is an answer about feasibility rather than about substance. The ethical answer, finally, is institutional, through committees, consent and registries; it leaves open the questions of justice, of collective consent, and of the financial dependence of field sites.

Two objections remain in the foreground. Transportability cannot be settled from within the apparatus, since it requires knowledge of auxiliary factors the trial does not produce (Deaton and Cartwright, 2018). And the selection of questions by the availability of an assignment mechanism has almost never been treated as a methodological problem, although it

determines what the discipline knows and what it does not. It is this last objection that the next section takes up.

6. The displacement of the object

Every objection reviewed in Section 5 was framed as a limit of validity. Taken together they in fact trace the outline of a new object. This section establishes the point along the three conditions set out in Section 2.

6.1 First condition: questions that have become peripheral

A question loses its centrality in a field when it can no longer be posed there without changing the format of an article. Four families of questions are in that position.

The first is the sectoral composition of output and its dynamics. Why an economy remains specialised in low-productivity activities, and by what mechanisms it moves out of them, is the historical question of the field. It is now handled in a literature no longer counted as experimental development economics and published elsewhere (Rodrik, 2016; Lin, 2011). The second family concerns international insertion: terms of trade, primary specialisation, reliance on external finance, debt sustainability. The third bears on the building of state capacity, understood as the capacity to administer, to tax and to implement, which Section 5.1 shows to determine the gap between the effect estimated and the effect obtained. The fourth gathers questions of distribution: the formation of rents, the structure of landholding, the position of economies within value chains. None of these questions has been refuted. No demonstration has established that they were badly posed. They appear to have become less central in the most visible segments of the field, and to have moved towards adjacent subfields, because the apparatus does not take them. Put this way the proposition asserts a tendency, and Section 9 measures how far the vocabulary of the field bears it out.

6.2 Second condition: selection by the assignment mechanism

That objects are selected by the availability of an identification design is visible in the structure of the literature itself. The best-covered domains, education, health, microfinance, cash transfers, plot-level agriculture, share one property: the intervention is divisible, attributable to an individual, a household, a school or a village, and observable over two or three years. The least-covered domains, exchange rate policy, banking

regulation, industrial policy, taxation of natural resources, trade negotiation, share the opposite property: the intervention is indivisible and applies to an economy as a whole.

This pattern is not the outcome of any ranking of stakes. It is a technical constraint converted into a research programme.

The geography of research lends the point indirect support. Porteous (2022) analyses 27,000 economics articles on Africa's 54 countries and finds that 93 per cent of the cross-country variation in research volume is explained by five variables: peacefulness, political institutions, international tourist arrivals, English as an official language, and population. Not one of these measures the economic weight of a country or the severity of the problems it faces. What they measure is the cost of access to the field for a foreign researcher. Within countries, research concentrates in higher-income regions and large urban areas. Porteous draws a conclusion this article adopts: since most research on African countries is context-specific, the absence of work on a country or a region means a thinner evidence base for the policymakers who work there. Madagascar, an island economy, francophone, classified among low-income countries, sits precisely in the part of the distribution that these five variables penalise. One reservation applies: Porteous's corpus covers all economics research on Africa and not experimental work alone, so it documents the geography of research in general rather than the selection specific to the assignment apparatus. It makes the latter plausible without isolating it.

The argument can be put simply. Research effort is allocated in proportion to identifiability multiplied by fundability, not in proportion to the magnitude of the phenomenon. So long as the two series coincide, the selection is harmless. Nothing guarantees that they coincide, and the case of economies with weak administrative capacity shows them diverging exactly where the stakes are heaviest.

6.3 Third condition: the requalification of statements

The third condition carries the heaviest consequences. The results of the apparatus are statements about programmes. They are routinely read, including by well-informed readers, as statements about development.

Microcredit is instructive here. Meager (2019) establishes that the average effect of microcredit expansions on household consumption and business activity is unlikely to be transformative. That is an exact statement about

seven evaluated programmes. It was received as a statement about the role of credit in the development of low-income economies, which it is not: none of the seven trials bears on the financial depth of an economy, on the structure of its banking system, or on the intersectoral allocation of credit. The question of development finance is distinct from the question of the average effect of a credit product on a marginal household, and the second supplies no answer to the first.

A rhetorical usage facilitates the requalification. Duflo (2017) offers the figure of the economist as plumber, attending to the details of implementation. The metaphor is apt and avowed. It describes a useful trade. It does not describe the historical object of the field, and its adoption as a disciplinary identity performs precisely the substitution this article describes.

6.4 A typology of the displacement

The displacement can be ordered along four dimensions, set out in Figure 2.

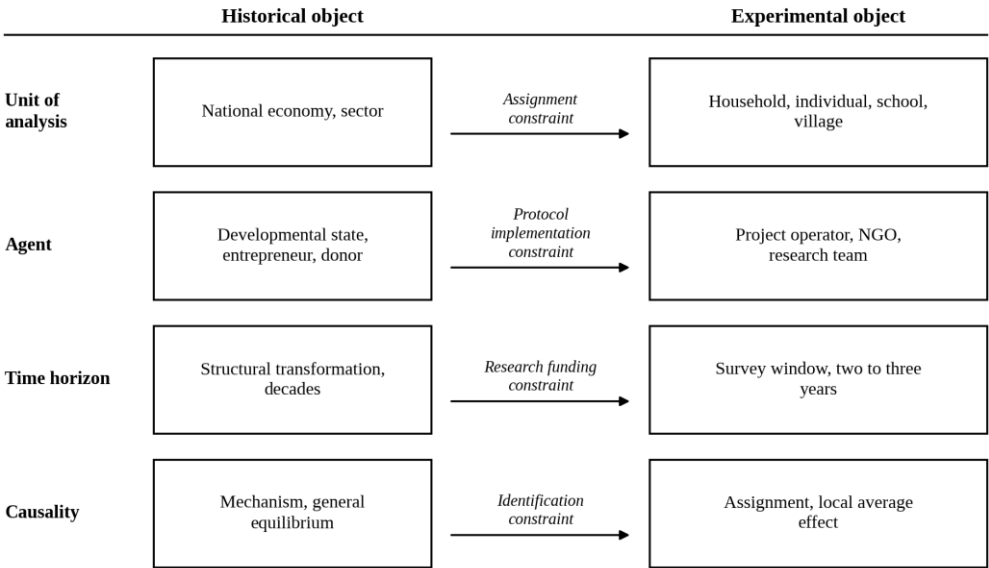


Figure 2. The displacement of the object of development economics: four dimensions

Source: author's elaboration.

Each row of Figure 2 reads the same way: what produced the shift was not a theoretical conviction but a requirement of the apparatus converted

into a definition of the object. The fourth row is the most consequential and calls for the most careful wording. Experimentation does not abolish the analysis of mechanisms: a great deal of experimental work studies them through mediators, intermediate outcomes, heterogeneity or behavioural models. What moves is the anchor of causal credibility, which now rests more on the procedure of assignment than on an explicit account of the mechanism through which the effect passes. The second row calls for the same care: the state is not absent from the experimental literature, it is repositioned, from central object of explanation to implementing institution, institutional constraint or contextual variable. Card, DellaVigna and Malmendier (2011) had already observed, of field experiments generally, that few are designed to test an explicit theoretical prediction. Basu (2014) draws the consequence: reasoned intuition, which is to say the exercise of theory, does not disappear from an experimental science; it becomes implicit there, and therefore uncontrolled.

6.5 The behavioural shift

One further observation completes the typology. Substituting the individual for the national economy as the unit of analysis opens the way to a behavioural account of poverty. Banerjee and Duflo (2011) organise their synthesis around the decisions of the poor, their informational constraints and their biases. The World Development Report 2015 (World Bank, 2015) institutionalises the reading. Kremer, Rao and Schilbach (2019) provide its reference statement in the Handbook of Behavioral Economics.

The corpus is of high quality and its results hold on their domain. The risk lies elsewhere. When the unit of analysis is the household and the mechanism is behavioural, the explanation of poverty migrates towards the decision properties of the poor, and the determinants that do not appear there, the structure of property, the price regime, the external constraint, administrative capacity, become parameters of context. Hoff and Stiglitz (2016) formulate the antidote from inside behavioural economics: behaviour is socially determined, which forbids treating structure as a backdrop. That such a reminder had to be issued is itself telling.

7. Recompositions: the object returns through the method

From the middle of the 2010s the experimental programme produced a series of corrections. The reading offered here is that each of them puts back an element of the displaced object, in a form the apparatus constrains. Figure 3 summarises the correspondence.

Correction since 2015		Element reinstated
Bayesian hierarchical aggregation <i>(Meager, 2019, 2022)</i>	→	The standing of a population of sites, against the single-site result
Bounds on external validity <i>(Andrews and Oster, 2019; Dehejia et al., 2021)</i>	→	Comparability of populations, hence structure
General equilibrium effects <i>(Egger et al., 2022)</i>	→	Prices, local markets, the multiplier
Experimentation at scale <i>(Muralidharan and Niehaus, 2017; Bold et al., 2018)</i>	→	Public administration as agent
Political economy of implementation <i>(Bédécarrats et al., 2019, 2020; Vivalt, 2020)</i>	→	Power relations between operator, state and population
Return of industrial policy <i>(Rodrik, 2016; Juhász et al., 2024)</i>	→	Productive structure and its transformation

Figure 3. Corrections devised since 2015 and the structural element each reinstates

Source: author's elaboration.

Aggregation. Meager (2019, 2022) treats results from multiple sites as draws from a distribution whose parameters, mean and dispersion, are to be estimated, and does so for the full distribution of effects rather than the mean alone. The approach settles part of the external validity debate by restating it statistically: the question is no longer "does this result travel" but "how large is between-site variance and what accounts for it". It nonetheless requires a population of comparable sites, that is, a theory of comparability that no single trial supplies.

Transportability. Andrews and Oster (2019) propose an approximation of external validity bias based on observable selection; Dehejia, Pop-Eleches

and Samii (2021) show, on a fertility natural experiment replicated across a large number of countries, that out-of-sample prediction is feasible provided one has structural covariates. Behind this work lies the uncomfortable result of Pritchett and Sandefur (2015): where context matters, a non-experimental estimate obtained in the right context can predict the local effect better than an experimental estimate obtained somewhere else. Applied mechanically, the hierarchy of evidence then yields the larger forecasting error.

Scale and agent. The accumulation of findings on scaling has made explicit what Section 6 states: the object evaluated and the object of public policy differ by the agent who executes. To acknowledge this is to make administrative capacity a variable rather than a parameter, which restores to the field a question it had left to political science and institutional economics.

The return of the macro. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, awarded in 2024 to Daron Acemoglu, Simon Johnson and James A. Robinson "for studies of how institutions are formed and affect prosperity" (Committee for the Prize in Economic Sciences, 2024), provides an institutional marker. Five years after the consecration of the experimental programme, the same body distinguished the historical and comparative programme. The award is a signal consistent with the reading defended here, namely the continued importance of comparative, historical and political-economic research; it does not by itself demonstrate any disciplinary reversal, and it would be imprudent to make it carry more.

The same movement is visible in industrial policy. Long treated as a disqualified object, it has been reconstituted as a legitimate field of study, with empirical results and a renewed analytical frame (Juhász, Lane and Rodrik, 2024). Rodrik (2016) documents besides the premature deindustrialisation of low and middle-income economies, a phenomenon that bears directly on the path of structural transformation on which earlier strategies rested. These are questions belonging to the historical object, treated with the tools available, without any experimental apparatus and without their scientific standing being contested.

What complementarity would require

Calls for methodological complementarity are frequent and rarely operational. Three requirements give the notion content.

The first concerns the order of operations. An experiment carries information for policy only if the question it addresses has been selected by an analysis of the structure of the problem, and not by the availability of an operator willing to randomise. This reverses the common practice, in which the choice of intervention precedes the analysis of the sector.

The second concerns the unit of assignment. Once the effect of interest passes through prices, local labour markets or public budgets, assignment must occur at the level where those mechanisms operate. Egger et al. (2022) provide both the demonstration and the cost. A practical implication follows for teams without such budgets: on these questions, administrative, customs and nationally representative survey data remain the only available apparatus, and their exploitation is not a second-best method.

The third concerns the standing of the implementing agent. Since the identity of the operator alters the effect (Vivalt, 2020; Bold et al., 2018), evaluation must treat administrative capacity as an explicit variable of the protocol, which means documenting the institutional conditions of implementation as demandingly as the characteristics of beneficiaries. Such documentation is at present the exception.

8. Discussion: an unsettled hierarchy

Three consequences follow.

The hierarchy of evidence and the hierarchy of questions do not coincide.

The first ranks methods by the quality of identification; the second should rank objects by the magnitude of what is at stake. The discipline built the first with rigour and let the second adjust to it. The practical consequence is measurable: on a set of questions that govern the living standards of several hundred million people, the stock of established knowledge is thin, not because the questions lack interest, but because they do not lend themselves to random assignment.

Internal validity is a property of execution, not of protocol. Young's (2019) results on inference, those of Bédécarrats et al. (2019) on data quality, and those of Ioannidis, Stanley and Doucouliagos (2017) on power all converge: the guarantee one believes to obtain from the assignment plan is in fact produced by the conduct of the survey, by effective take-up, by the handling of contamination and by the statistical treatment. These elements are rarely documented in the detail required,

and the distance between the promise of the protocol and the reality of execution remains the blind spot of the programme.

Each correction reproduces the constraint. Every solution to the limits of the apparatus demands more resources: more sites for aggregation, whole markets for general equilibrium, agreements with national administrations for experimentation at scale. The consequence is a growing concentration of evidence production in the hands of a small number of teams and funders. A method that was to democratise proof, since it asked only for a clean protocol and a field site, ends by producing an economy of research more concentrated than the one it replaced. That is a finding in the political economy of knowledge, and it bears first on the universities of the countries where the trials take place.

This last observation has a direct bearing on research conducted from African universities, and it deserves to be spelled out. Three consequences follow from the mechanism described in this article, and they compound.

The first is compositional. What an African policymaker can consult, when asking what is known about a problem in their own country, is a body of evidence made almost entirely of programme-level effects. The questions on which macroeconomic and industrial choices are actually decided, the exchange rate regime, tariff structure, the taxation of extractive rents, the sequencing of industrial support, debt sustainability, are the ones the assignment apparatus cannot take, and are therefore the least covered, as Section 6.2 sets out.

The second is geographical. Porteous (2022) shows that the distribution of economics research across African countries follows the cost of access to the field rather than the weight of the problems, so the thin evidence base is thinnest precisely where states are weakest and stakes highest. Madagascar sits in that part of the distribution.

The third bears on how research is funded, and here the claim must be stated as a mechanism rather than as a measured fact, since no source disaggregates research funding by object of study. Evaluation funding is attached to interventions: it is disbursed by agencies and foundations that need to know whether a programme worked, on an horizon compatible with a project cycle. Structural analysis has no equivalent commissioning channel. Where the marginal research budget available to an African university department comes through that channel, the composition of

what gets studied follows the composition of what gets commissioned, and the displacement described here is reproduced locally, at the level of individual careers. The position of field supplier, with subordinate access to the data produced locally, is the default arrangement that results. Leaving it behind means choosing objects whose analysis does not depend on an experimental design: the structure of domestic prices, the budgetary effects of subsidy regimes, margin formation along export chains, the collection capacity of tax administrations. These are questions that administrative, customs and survey data allow one to treat, and whose value does not depend on access to trial funding.

9. Measuring the reweighting: a bibliometric probe

The preceding sections rest on an interpretation. This one tests part of it. If the object of the field was reweighted, the vocabulary of its journals should record the change, since a research programme leaves its trace in the words its articles use to state what they are about.

9.1 Data and method

The corpus is drawn from OpenAlex (Priem, Piwowar and Orr, 2022), queried on 2 September 2026. Seven journals were selected because they carry the field's mainstream and its heterodox margins alike: *Journal of Development Economics* (ISSN 0304-3878), *World Development* (0305-750X), *Journal of Development Studies* (0022-0388), *Economic Development and Cultural Change* (0013-0079), *World Bank Economic Review* (0258-6770), *Journal of African Economies* (0963-8024) and *Oxford Development Studies* (1360-0818). Records are restricted to items typed as articles and published between 1990 and 2025, which yields 18,592 articles.

Two lexical families are counted in titles and abstracts. The experimental family is *randomized*, *randomised*, *randomization*, *randomisation*. The structural family is *structural transformation*, *structural change*, *industrial policy*, *state capacity*, *terms of trade*, searched as phrases. The denominator in each year is the total number of articles published by the same journals in that year, so what is reported is a share and not a count, which neutralises the growth in journal output. As a robustness check the experimental family was also searched in titles alone.

Four limits must be stated before the results. A lexical family is a proxy for an object, not a content analysis: an article may study structural

transformation without using the phrase. Abstract coverage in OpenAlex is uneven before 2000, which depresses the early counts of both families; since the experimental family is the one that grows, this bias understates rather than inflates the change measured here. The 2025 volume is still being indexed. And neither family is exhaustive, so the levels matter less than the trajectories and the ratio between them.

9.2 Results

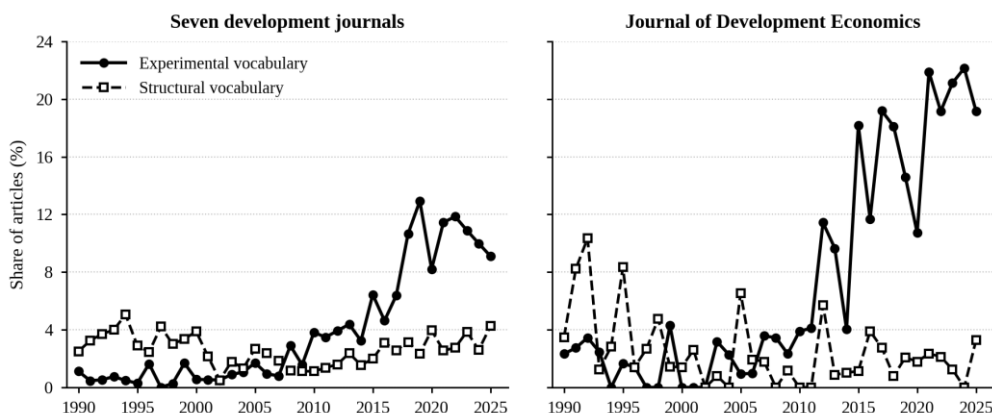


Figure 4. Share of articles carrying the experimental and the structural vocabulary, seven development journals and the Journal of Development Economics, 1990-2025

Source: OpenAlex; author's calculations.

Pooled across the seven journals, the experimental share rises from 0.71 per cent of articles in 1990-1999 to 1.21 per cent in 2000-2009, 4.23 per cent in 2010-2015 and 9.63 per cent in 2016-2025, with a peak of 12.9 per cent in 2019. Between the first and the last of these periods the share is multiplied by 13.5. The title-only check gives the same picture in starker terms: one article in the whole of the 1990s carries the vocabulary in its title, against 96 in 2016-2025.

The structural share does not mirror this rise. It stands at 3.47 per cent in 1990-1999, falls to 1.84 per cent in 2000-2009 and 1.68 per cent in 2010-2015, then recovers to 3.12 per cent in 2016-2025. Its trajectory is a U, not a decline. The ratio of the two families tells the story more compactly: 0.20 in the 1990s, 0.66 in the 2000s, 2.52 in 2010-2015, 3.08 in 2016-2025.

The *Journal of Development Economics*, taken alone, shows the movement in its purest form. The experimental share there goes from 1.84 per cent in 1990-1999 to 1.73 per cent in 2000-2009, 8.64 per cent in 2010-2015 and 18.36 per cent in 2016-2025, reaching 19.2 per cent in 2025: close to one article in five. The structural share in the same journal falls from 4.24 per cent to 1.73 per cent and never returns above 2 per cent.

9.3 What the probe establishes, and what it does not

Three findings follow, and the second is not the one the argument of Sections 1 to 6 would have predicted.

The reweighting is real and its magnitude is large. On the field's flagship journal it is not a matter of emphasis: the experimental vocabulary moves from the margin of the journal to a fifth of its output within a generation, and no other methodological vocabulary in the discipline has travelled that distance in that time.

The structural vocabulary did not disappear. It contracted sharply during the 2000s, the decade in which the experimental programme installed itself, and it recovered afterwards. This is the recomposition described in Section 7, now measured rather than asserted, and it counts against the strong version of the displacement thesis. The version this article defends, a reweighting rather than an elimination, is the version the data support.

The recovery is uneven, and where it is absent is instructive. It is visible in the pooled series but not in the *Journal of Development Economics*, whose structural share stays near 2 per cent throughout the recovery period. Since that journal is the outlet where the discipline polices its own boundary most strictly, the plausible reading is that structural questions returned mainly through the multidisciplinary journals of the pool. A per-journal decomposition would be needed to confirm it, and that is the natural next step of this measurement.

What the probe does not establish should be equally clear. It measures words, not objects. It cannot show that a question was abandoned, only that the vocabulary in which it was once posed became rarer while another became common. The three conditions of Section 2 are therefore met unequally: the second and the third rest on the argument developed in Section 6, while the first now has a quantitative anchor for the period 1990-2025. That is an improvement on the interpretive claim this article began with, and it is not a proof.

10. Conclusion

The experimental turn equipped development economics with an apparatus of identification it had lacked, and with a discipline of evidence whose benefits extend beyond its domain of application. None of that is contested here. The thesis defended rests on three propositions. The experimental turn altered the field's dominant unit of analysis. It favoured the study of programmes over the study of structures, without removing the latter from economic research. And the recent corrections progressively reinstate markets, aggregation, institutions and the state. The three propositions form an interpretation to be tested rather than an established fact, and the reweighting they describe was carried out by constraints of design rather than by a theoretical argument.

The period opened in the mid-2010s is one of partial return. Aggregation restores the population of sites, general equilibrium restores markets, experimentation at scale restores the state, and industrial policy has become a legitimate object once more. The return proceeds under methodological constraint, which limits its reach but secures its solidity.

Two limitations must be stated plainly. The measurement of Section 9 counts words rather than objects, so it establishes that the vocabulary of the field was reweighted and leaves the reweighting of its questions an inference, well supported but an inference still. And the probe covers seven journals: a per-journal decomposition, and an extension to political economy and growth journals, would show whether the structural vocabulary returned to development economics or migrated out of it. That extension is the natural continuation of this work.

What is at stake exceeds the sociology of a discipline. A development economics that takes as its object only what it can identify cleanly leaves untreated the determinants of the very gaps in living standards it exists to explain. The point on which the field must pronounce is the criterion by which its questions are selected, not the quality of its answers.

References

Acemoglu, D. (2010). Theory, general equilibrium, and political economy in development economics. *Journal of Economic Perspectives*, 24(3), 17-32. <https://doi.org/10.1257/jep.24.3.17>

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation.

- American Economic Review, 91(5), 1369-1401.
<https://doi.org/10.1257/aer.91.5.1369>
- Allcott, H. (2015). Site selection bias in program evaluation. *Quarterly Journal of Economics*, 130(3), 1117-1165.
<https://doi.org/10.1093/qje/qjv015>
- Andrews, I., & Oster, E. (2019). A simple approximation for evaluating external validity bias. *Economics Letters*, 178, 58-62.
<https://doi.org/10.1016/j.econlet.2019.02.005>
- Angrist, J., Azoulay, P., Ellison, G., Hill, R., & Lu, S. F. (2017). Economic research evolves: Fields and styles. *American Economic Review*, 107(5), 293-297. <https://doi.org/10.1257/aer.p20171117>
- Banerjee, A. V. (2020). Field experiments and the practice of economics. *American Economic Review*, 110(7), 1937-1951.
<https://doi.org/10.1257/aer.110.7.1937>
- Banerjee, A. V., & Duflo, E. (2009). The experimental approach to development economics. *Annual Review of Economics*, 1, 151-178.
<https://doi.org/10.1146/annurev.economics.050708.143235>
- Banerjee, A. V., & Duflo, E. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: PublicAffairs. ISBN 978-1-58648-798-0
- Banerjee, A. V., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives*, 31(4), 73-102.
<https://doi.org/10.1257/jep.31.4.73>
- Banerjee, A. V., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., & Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, 348(6236), 1260-1269.
<https://doi.org/10.1126/science.1260799>
- Barrett, C. B., & Carter, M. R. (2010). The power and pitfalls of experiments in development economics: Some non-random reflections. *Applied Economic Perspectives and Policy*, 32(4), 515-548.
<https://doi.org/10.1093/aep/ppq023>
- Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106(2), 407-443.
<https://doi.org/10.2307/2937943>

Basu, K. (2014). Randomisation, causality and the role of reasoned intuition. *Oxford Development Studies*, 42(4), 455-472.
<https://doi.org/10.1080/13600818.2014.961414>

Bédécarrats, F., Guérin, I., & Roubaud, F. (2019). All that glitters is not gold: The political economy of randomized evaluations in development. *Development and Change*, 50(3), 735-762.
<https://doi.org/10.1111/dech.12378>

Bédécarrats, F., Guérin, I., & Roubaud, F. (eds) (2020). *Randomized Control Trials in the Field of Development: A Critical Perspective*. Oxford: Oxford University Press. ISBN 978-0-19-886536-0.
<https://doi.org/10.1093/oso/9780198865360.001.0001>

Bédécarrats, F., Guérin, I., Morvant-Roux, S., & Roubaud, F. (2019). Estimating microcredit impact with low take-up, contamination and inconsistent data: A replication study of Crépon, Devoto, Duflo and Parienté (*American Economic Journal: Applied Economics*, 2015). *International Journal for Re-Views in Empirical Economics*, 3(2019-3).
<https://doi.org/10.18718/81781.12>

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2018). Experimental evidence on scaling up education reforms in Kenya. *Journal of Public Economics*, 168, 1-20.
<https://doi.org/10.1016/j.jpubeco.2018.08.007>

Cameron, D. B., Mishra, A., & Brown, A. N. (2016). The growth of impact evaluation for international development: How much have we learned? *Journal of Development Effectiveness*, 8(1), 1-21.
<https://doi.org/10.1080/19439342.2015.1034156>

Card, D., DellaVigna, S., & Malmendier, U. (2011). The role of theory in field experiments. *Journal of Economic Perspectives*, 25(3), 39-62. <https://doi.org/10.1257/jep.25.3.39>

Cartwright, N., & Hardie, J. (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press. ISBN 978-0-19-984160-8

Christensen, G., & Miguel, E. (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, 56(3), 920-980.
<https://doi.org/10.1257/jel.20171350>

Committee for the Prize in Economic Sciences in Memory of Alfred Nobel (2024). *The Sveriges Riksbank Prize in Economic Sciences in*

Memory of Alfred Nobel 2024: Daron Acemoglu, Simon Johnson, James A. Robinson, "for studies of how institutions are formed and affect prosperity". Stockholm: Royal Swedish Academy of Sciences. <https://www.nobelprize.org/prizes/economic-sciences/2024/summary/> (accessed 2 September 2026)

Crépon, B., Devoto, F., Duflo, E., & Parienté, W. (2015). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco. *American Economic Journal: Applied Economics*, 7(1), 123-150. <https://doi.org/10.1257/app.20130535>

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48(2), 424-455. <https://doi.org/10.1257/jel.48.2.424>

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2-21. <https://doi.org/10.1016/j.socscimed.2017.12.005>

Dehejia, R., Pop-Eleches, C., & Samii, C. (2021). From local to global: External validity in a fertility natural experiment. *Journal of Business & Economic Statistics*, 39(1), 217-243. <https://doi.org/10.1080/07350015.2019.1639407>

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4), 795-813. <https://doi.org/10.1257/aer.91.4.795>

Duflo, E. (2017). Richard T. Ely Lecture: The economist as plumber. *American Economic Review*, 107(5), 1-26. <https://doi.org/10.1257/aer.p20171153>

Duflo, E. (2020). Field experiments and the practice of policy. *American Economic Review*, 110(7), 1952-1973. <https://doi.org/10.1257/aer.110.7.1952>

Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. In T. P. Schultz & J. A. Strauss (eds), *Handbook of Development Economics* (vol. 4, ch. 61, pp. 3895-3962). Amsterdam: Elsevier. [https://doi.org/10.1016/S1573-4471\(07\)04061-2](https://doi.org/10.1016/S1573-4471(07)04061-2)

Durlauf, S. N., Johnson, P. A., & Temple, J. R. W. (2005). Growth econometrics. In P. Aghion & S. N. Durlauf (eds), *Handbook of*

Economic Growth (vol. 1A, ch. 8, pp. 555-677). Amsterdam: Elsevier.
[https://doi.org/10.1016/S1574-0684\(05\)01008-7](https://doi.org/10.1016/S1574-0684(05)01008-7)

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., & Walker, M. (2022). General equilibrium effects of cash transfers: Experimental evidence from Kenya. *Econometrica*, 90(6), 2603-2643.
<https://doi.org/10.3982/ECTA17945>

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives*, 9(2), 85-110.
<https://doi.org/10.1257/jep.9.2.85>

Hoff, K., & Stiglitz, J. E. (2016). Striving for balance in economics: Towards a theory of the social determination of behavior. *Journal of Economic Behavior & Organization*, 126(B), 25-57.
<https://doi.org/10.1016/j.jebo.2016.01.005>

Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature*, 48(2), 399-423. <https://doi.org/10.1257/jel.48.2.399>

Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, C. (2017). The power of bias in economics research. *Economic Journal*, 127(605), F236-F265. <https://doi.org/10.1111/ecoj.12461>

Jatteau, A. (2013). Expérimenter le développement ? Des économistes et leurs terrains. *Genèses*, 93(4), 8-28.
<https://doi.org/10.3917/gen.093.0008>

J-PAL (2026). By the Numbers. Abdul Latif Jameel Poverty Action Lab, Massachusetts Institute of Technology.
<https://www.povertyactionlab.org/2025/by-the-numbers> (accessed 1 September 2026)

Juhász, R., Lane, N., & Rodrik, D. (2024). The new economics of industrial policy. *Annual Review of Economics*, 16, 213-242.
<https://doi.org/10.1146/annurev-economics-081023-024638>

Kremer, M. (2020). Experimentation, innovation, and economics. *American Economic Review*, 110(7), 1974-1994.
<https://doi.org/10.1257/aer.110.7.1974>

Kremer, M., Rao, G., & Schilbach, F. (2019). Behavioral development economics. In B. D. Bernheim, S. DellaVigna & D. Laibson (eds), *Handbook of Behavioral Economics: Applications and Foundations* (vol. 2, ch. 5, pp. 345-458). Amsterdam: Elsevier.
<https://doi.org/10.1016/bs.hesbe.2018.12.002>

Krugman, P. (1992). Toward a counter-counterrevolution in development theory. *World Bank Economic Review*, 6(suppl. 1), 15-38. https://doi.org/10.1093/wber/6.suppl_1.15

Labrousse, A. (2010). Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d'un outil de preuve et de gouvernement. *Revue de la régulation*, 7. <https://doi.org/10.4000/regulation.7818>

Labrousse, A. (2016). Not by technique alone: A methodological comparison of development analysis with Esther Duflo and Elinor Ostrom. *Journal of Institutional Economics*, 12(2), 277-303. <https://doi.org/10.1017/S1744137415000415>

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The Manchester School*, 22(2), 139-191. <https://doi.org/10.1111/j.1467-9957.1954.tb00021.x>

Lin, J. Y. (2011). New structural economics: A framework for rethinking development. *World Bank Research Observer*, 26(2), 193-221. <https://doi.org/10.1093/wbro/lkr007>

Meager, R. (2019). Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments. *American Economic Journal: Applied Economics*, 11(1), 57-91. <https://doi.org/10.1257/app.20170299>

Meager, R. (2022). Aggregating distributional treatment effects: A Bayesian hierarchical analysis of the microcredit literature. *American Economic Review*, 112(6), 1818-1847. <https://doi.org/10.1257/aer.20181811>

Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1), 159-217. <https://doi.org/10.1111/j.1468-0262.2004.00481.x>

Muralidharan, K., & Niehaus, P. (2017). Experimentation at scale. *Journal of Economic Perspectives*, 31(4), 103-124. <https://doi.org/10.1257/jep.31.4.103>

Porteous, O. (2022). Research deserts and oases: Evidence from 27 thousand economics journal articles on Africa. *Oxford Bulletin of Economics and Statistics*, 84(6), 1235-1258. <https://doi.org/10.1111/obes.12510>

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv preprint. <https://doi.org/10.48550/arXiv.2205.01833>

Pritchett, L., & Sandefur, J. (2015). Learning from experiments when context matters. *American Economic Review*, 105(5), 471-475. <https://doi.org/10.1257/aer.p20151016>

Ravallion, M. (2020). Should the randomistas (continue to) rule? In F. Bédécarrats, I. Guérin & F. Roubaud (eds), *Randomized Control Trials in the Field of Development: A Critical Perspective* (ch. 3). Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198865360.003.0003>

Rodríguez, F., & Rodrik, D. (2000). Trade policy and economic growth: A skeptic's guide to the cross-national evidence. *NBER Macroeconomics Annual*, 15, 261-325. <https://doi.org/10.1086/654419>

Rodrik, D. (2016). Premature deindustrialization. *Journal of Economic Growth*, 21(1), 1-33. <https://doi.org/10.1007/s10887-015-9122-3>

Rosenstein-Rodan, P. N. (1943). Problems of industrialisation of Eastern and South-Eastern Europe. *Economic Journal*, 53(210/211), 202-211. <https://doi.org/10.2307/2226317>

Sen, A. (1983). Development: Which way now? *Economic Journal*, 93(372), 745-762. <https://doi.org/10.2307/2232744>

Vivalt, E. (2020). How much can we generalize from impact evaluations? *Journal of the European Economic Association*, 18(6), 3045-3089. <https://doi.org/10.1093/jeea/jvaa019>

World Bank (2015). *World Development Report 2015: Mind, Society, and Behavior*. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-0342-0>

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics*, 134(2), 557-598. <https://doi.org/10.1093/qje/qjy029>