

Uncovering Hidden Configurations: Specifying Outcomes in National Innovation Systems with Archetypal Analysis

Research Background

In comparative social science, certain cases anchor our theoretical imagination: Germany as the archetypal coordinated market economy, the United States as the liberal market ideal type, Japan as the developmental state (Hall and Soskice, 2001; Johnson, 1982). These theoretically derived types structure how configurational analysis proceeds—they provide the categories into which cases are sorted and the outcomes whose causes are investigated. The same cases also dominate the study of national innovation systems, where they serve as implicit benchmarks against which other countries are compared. Yet our empirical analysis of 147 countries across eight years reveals a different landscape. Alongside these familiar cases, we find persistent configurations that no major theoretical tradition predicts: Romania and Latvia, among others, appear as stable reference points year after year, anchoring the global innovation space in ways that existing typologies cannot account for.

This observation points to a methodological question: where do the ideal types that underpin configurational analysis come from? Are they derived from theoretical tradition alone, or can they be grounded empirically? Configurational methods—from QCA to Bayesian Rule Sets—take types as given; they do not ask whether those types adequately capture the full range of configurations present in the data, including those that theory has overlooked. Addressing this requires a method that derives types inductively—a tool for identifying empirical analogues to Weberian ideal types before asking what leads to them.

Research Gap

Configurational analysis begins with types. Whether drawn from the varieties of capitalism literature, national innovation systems frameworks, or typologies of developmental states, these types serve as the foundation for subsequent inquiry. The empirical question then becomes: what conditions lead cases to fall into one type rather than another? This is a productive research strategy. What it leaves unexamined, however, is whether the types derived from theory are the types that actually exist in the data. Configurational methods, including QCA, take the specification of types as their point of departure; they are not designed to ask whether those types are empirically adequate representations of the phenomena under study.

This gap is conceptual rather than technical. Weber's conception of ideal types serves two distinct purposes: as tools for causal explanation and as pure forms whose mixtures constitute empirical reality. QCA has operationalized the first purpose with considerable sophistication. The second—systematically deriving ideal types from empirical observation in dialogue with theoretical tradition—has until now lacked a method capable of realizing it. Without such a method, we risk building causal explanations on typological foundations that may be incomplete or misspecified. The question of "why" can only be as sound as our understanding of "what." This paper introduces Archetypal Analysis to fill that gap.

Research Questions

This study asks three interrelated questions. First, what are the empirically grounded pure types—or "archetypes"—of national innovation systems that actually exist in the data, beyond those specified by existing theoretical frameworks? Second, how do actual

countries mix these pure types, and what patterns of hybridity emerge when we allow types to be derived empirically rather than imposed a priori? Third, what can the long-term dynamics of these archetypes reveal about the evolution of global innovation landscapes, and how might this empirical grounding of types inform subsequent configurational analysis?

By answering these questions, we demonstrate a method that makes possible what configurational approaches have long presupposed: a systematic way of establishing what the relevant types are before asking what leads to them.

A Methodological Innovation

To answer these questions, we introduce Archetypal Analysis (AA) to comparative sociology. AA is an unsupervised learning method designed to identify extreme configurations at the boundaries of data and represent each observation as a convex combination of these archetypes via α -coefficients. Unlike PCA, which reduces data to average dimensions, or cluster analysis, which forces cases into discrete categories, AA serves two functions that configurational methods cannot perform for themselves: it lets relevant types emerge from the data rather than being imposed a priori, and it quantifies the degree to which each case partakes of multiple types simultaneously—capturing the hybrid character of empirical reality that typological traditions have long recognized but never systematically measured.

We further extend this approach with dynamic convex hull analysis, tracking the full trajectory of innovation archetypes across eight years (2013-2020) for 147 countries using 73 indicators from the Global Innovation Index. Hungarian algorithm matching ensures optimal correspondence of archetypes across years, enabling the first systematic tracking of archetype lifecycles in comparative research.

Data and Key Findings

Our analysis yields four discoveries.

First, we identify 44 distinct archetypes across the eight-year period, with 16 countries persisting as permanent archetypes throughout. These include not only familiar cases—United States, United Kingdom, Japan, China, Switzerland, Sweden—but also overlooked ones: Romania, Latvia, Iceland, Luxembourg, New Zealand, Thailand, Mongolia, Kyrgyzstan, Niger, and Yemen. This finding reveals that "high innovation" is not a unitary phenomenon but encompasses multiple co-existing innovation modes. It reveals a global landscape of multiple, co-existing innovation modes. The evidence supports that national innovation systems exhibit configurational heterogeneity – different ways of combining institutional, human capital, market, and creative dimensions – that aggregate metrics obscure.

Second, the global innovation space is shrinking. Convex hull volume decreased by 8.5% over the period (from 871.9 to 797.6), while compactness declined by 12.2% (from 0.323 to 0.284). These geometric measures indicate that countries' innovation profiles are converging—the extremes are becoming less extreme, and the overall diversity of innovation configurations is contracting. This suggests a global tendency toward a common innovation model, with implications for both theory and policy.

Third, this study finds that archetypes themselves are remarkably stable despite landscape convergence. Average annual matching success of 96.0% demonstrates that while the overall space contracts, the positions of specific archetypes remain coherent and trackable. This stability validates the archetype construct as a meaningful unit of analysis and opens possibilities for studying archetype-specific dynamics—which archetypes are

gaining or losing members, which are moving in the PC space, and what factors drive these changes.

Fourth, the co-existence of multiple permanent archetypes with differing innovation profiles. The USA, Sweden, and China, for instance, each occupy distinct vertices of the convex hull, reflecting qualitatively different configurations across institutional, human capital, market, and creative dimensions. Their differences are not scalar but configurational: they represent distinct ways of combining these dimensions, even as aggregate scores place them in the same "high innovation" category.

Contributions

This paper contributes in two ways.

Methodologically, we introduce a systematic approach for identifying empirical analogues of Weberian ideal types and quantifying their mixture in real cases—a task that configurational methods have long presupposed but could not perform for themselves. By letting relevant types emerge from the data rather than imposing them a priori, and by capturing each case's hybrid character through α -coefficients, this approach lays the groundwork for more nuanced causal analysis. It ensures that the outcomes of interest are properly specified as configurations before questions of causation are asked. For scholars working with configurational methods (e.g., QCA) and, more broadly, for those interested in ideal type analysis across the social sciences, this provides a tool to operationalize concepts that have long remained at the level of theoretical intuition.

Empirically, we provide the first systematic mapping of global innovation diversity and its dynamics. The identification of 16 permanent archetypes—including both familiar and overlooked cases—reveals a landscape of persistent heterogeneity beneath the trend toward convergence. The documentation of shrinking convex hull volume quantifies, for the first time, the extent to which national innovation systems are becoming more alike. These findings open new avenues for investigating the conditions that lead countries toward different innovation modes, and for understanding why some archetypes persist while others emerge or disappear. For scholars of science and technology innovation, this provides a new lens for seeing both the forest—global trends toward convergence—and the trees—the specific countries that anchor the extremes of innovation space.

Keywords: Archetypal Analysis, Configurational Theory, Weberian Typology, National Innovation Systems, Comparative Methods

Reference:

Hall, P. A., & Soskice, D. (2001). *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford University Press.

Johnson, C. (1982). *MITI and the Japanese miracle: the growth of industrial policy, 1925-1975*. Stanford university press.