Modern China
2021, Vol. 47(1) 3–25

DOI: 10.1177/0097700420924603
journals.sagepub.com/home/mcx
SAGE

Rethinking the Formalism-Substantivism Debate in Social Science: A Perspective from Recent Developments in Economic Methodology*

Yuan Gao[1]

Abstract

Taking as its point of departure the 1960s formalism-substantivism debate in social science methodology, this article argues that what is distinctive about the new development of formalism in economics since then is mainly the prevalence of using "complete models"—tractable, manipulable, and fully specified mathematical objects—to construct and express theories. The objective of complete models is not to establish general laws, but to formulate auxiliary devices of cognition to facilitate the explanation of targeted aspects of the empirical world; not to create idealistic or ideological discourses, but to derive implications with empirically delimited utility—this in order to make inferences that cannot be achieved via purely qualitative methods. This methodological trend is to some extent a substantivization of formalist economics. Exploring its nature can help clarify the unique cognitive value

[1]Renmin University of China, Beijing, China

*Editor's note: This article was presented at the conference on "The Social Science of Practice and China Research" in honor of Professor Philip C. C. Huang's 80th year (held at the School of Sociology of Renmin University in Beijing). The articles are being published in different groups in *Open Times* 开放时代, *Rural China-Zhongguo xiangcun yanjiu* 中国乡村研究, and *Modern China*, a total of 20 articles.

Corresponding Author:
Yuan Gao, School of Agricultural Economics and Rural Development, Renmin University of China, 50 Zhongguancun Street, Beijing 100872, China.
Email: hlbbk@163.com

of contemporary formalism and answer the question of why substantivism is still an irreplaceable approach to social scientific studies, even in an age dominated by formalism.

Keywords
formalism, substantivism, complete models, empirically delimited utility, economic methodology

The 1960s saw a famous formalist-substantivist debate over economic methodology, turning mostly on criticisms of neoclassical economics by a group of economic anthropologists and economic historians headed by Karl Polanyi.[1] These critics argued that neoclassical economics was mistaken in moving away from the substantive content of the economy, such as particular modes of production, consumption, trading, and the distribution of the surplus of a society during a certain historical period, to the abstract, and in its equating the economy with a mere collection of generally conceptualized activities, such as making "rational choices" under the constraints of resource scarcity. To replace this "formalist" approach of neoclassical economics, they advocated a "substantivist" approach, emphasizing the empirical reality of material production and nonmarket institutions rather than the abstraction and formalization of general patterns of human behavior. This debate has continued to influence discussions on methodology in the social sciences, especially in the subfields of economic anthropology and economic history, as well as in the sphere of area studies, which is concerned more with the particularities of specific countries and regions than the universal rules governing social systems.[2]

Today, both the methodological paradigms and the contents of the social sciences are quite different from those in the 1960s. Remarkably, there are few voices in mainstream economics that speak on behalf of substantivism.[3] Formalism has simply overwhelmed the practice of economic research. And formalism itself has also changed in many respects and has revealed characteristics that were absent in the 1960s. First, the dominant way of doing economic theory now is to build "complete models," that is, tractable but highly simplified models with fully specified settings, and deduce implications from them with mathematical rigidity.[4] Theoretical statements in economics now look just like theorems and proofs in mathematical articles, in sharp contrast to the formalist style in the 1960s where mathematics in many instances still performed an illustrative function, and there was no felt necessity to structure one's entire analysis along the lines of a strict "definition-theorem-proof" format. Second, some prominent economic theorists have clearly realized

that economic theories in terms of complete models cannot be readily equated with descriptions of real economic systems, and that the cognitive function of these models is not to predict real economic consequences in an exact manner à la natural science. Complete models, in their view, are most valuable when they can perform as a sort of "theoretical case," showing how things could be in a "theoretical world" that is definitely not the empirical world we directly observe (Gilboa et al., 2014). By contrasting theoretical cases with empirical cases in real economic systems, we can sharpen our understanding of the mechanisms or dynamics behind real-world cases. As for these new methodological trends in mainstream economics, how should one interpret their meaning and link them to the formalism-substantivism distinction which has long stimulated reflections on the philosophy of social science?

Apart from these new trends in social science, human society itself has, needless to say, also seen significant changes since the 1960s. Big data and formalized mathematical models are exerting ever more influence in guiding and organizing socioeconomic activities. In this situation, how should one evaluate the importance of substantivism, which argues for a more qualitative and less formalized style of research? What benefit can be derived from a dialogue between substantivism and formalism at a time when the latter is hegemonic in both social scientific studies and socioeconomic practice?

With these questions in mind, this article revisits the old formalism-substantivism debate, not to argue again for one or the other of the binary choices between the two approaches, but to delineate a new standpoint from which one can utilize and integrate insights from both sides. The first step of our revisit is to highlight the most remarkable feature that distinguishes formalism today from that of the 1960s: the use of complete models as a standard procedure for developing and expressing economic theories. We must emphasize that the core of the 1960s formalism-substantivism debate—which approach is better at representing the essence of real economic activities—is losing its significance. This is because economic theorists today are increasingly aware that the economic models they use are by no means accurate reflections of the real world but only idealistic constructs in a theoretical/hypothetical world. This fact points to the value of contemporary substantivism—it functions in a realm where formalized models are disconnected from, or, unable to represent and conceptualize, the empirical/real world.

The methodological value of a complete model lies precisely in the fact that it is simplified and structuralized, in sharp contrast to the anti-reductionist complexity of the empirical world. Only in this way can it be "tractable." That is, it makes it possible to solve the model explicitly and derive useful implications from it in a mathematically feasible and rigorous manner. Furthermore, thanks to the complete specification of the model settings, one

can manipulate its parameters to see what would happen if certain postulations were satisfied or changed, and thus one can "observe" the causal processes or mechanisms in a purely theoretical way. Using these theoretical observations as auxiliary knowledge, one can see the mechanisms and logic behind the real-world case under consideration—for which direct in-case observation would not reveal more, and controlled experiments would be difficult, if not impossible, to perform—in order to make valid causal inferences. By the same token, complete models are not intended to provide general laws to cover as many empirical facts as possible, but as localized theoretical devices with empirically delimited utility.

Finally, to make substantivism a vigorous ingredient of the philosophy of social science today requires underlining its pragmatic meaning, that is, a methodology that seeks to make explicit the actual nature of empirical cases in the real world. In contrast to this pragmatic approach, the traditional concept of substantivism focuses on its substantive difference from formalism over the connotations of the economic, and emphasizes the materiality of the economic and the prevalence of nonmarketized modes of institutions, as opposed to the formalist conceptualization of the economic as a collection of abstract, universal patterns of means-ends choice behavior, usually around marketized institutions. The pragmatic interpretation, in our view, could help substantivism go beyond its past realms of economic anthropology and economic history, to become a more applicable concept for general methodological discussion. With this in mind, this article also proposes the phrase "substantivization of formalist economics" to capture the essence of the above-mentioned methodological development of mainstream economics, and to dispel the myth of equating economics with a science to establish general laws prescribing how the empirical world works.

The New Trend of Formalism: Using Complete Models to Articulate Theories

Contemporary Formalism and Complete Models versus the Old Formalism

Formalism looks impressive, with its mathematical symbols, formulas, equations, and so on, in contrast to substantivist theories, which are mainly written in natural language, with fewer mathematized ingredients. The mathematization of formalism grew out of an attempt to mimic the natural sciences, especially physics. Since the age of Galileo-Newton, physics has been increasingly engaged in a research paradigm that combines the effort to develop physical laws in terms of mathematical formulas and to

examine these laws through controlled experiments.[5] Though it is difficult to conduct experiments on economic systems, the idea of using a formalized language like mathematics to express economic theories was evident as early as in classical political economy and even before that in the work of the physiocrats. In the eighteenth century, François Quesnay used numeric tables to depict how economic surplus is transferred from one social class to another. In the early nineteenth century, David Ricardo engaged in arithmetic exercises in an agricultural production model to show how wages, rents, and profits are formed. Half a century later, Karl Marx tried to use algebraic equations to capture the long-term dynamics of the capitalist economy. And in the last quarter of the nineteenth century, along with the so-called marginal revolution, calculus was introduced into economics to show how consumers and firms behave facing price incentives.[6] Until the 1930s, when John Maynard Keynes' pathbreaking work laid the foundation of modern macroeconomics, the mathematical tools of calculus and linear algebra, also the subjects of the core courses today for non-mathematics majors in science and engineering departments, had become a standard requirement for education in economics and for research in the English-speaking academy. Economics articles and books up to that point had developed a style that mixed both verbal arguments and mathematical analysis.

The formalism of the 1960s inherited these waves of the mathematization of economic studies. However, it is unlike the formalism of today in three respects, even though both iterations are marked by mathematization. First, the major tools of analysis of the formalists of the 1960s were incompletely specified models—that is, ones where model settings were only broadly indicated and flexibility and ambiguity were tolerated. Second, their publications were largely verbal. While mathematical expressions were embedded in their writings, it was not in a rigidly structuralized manner. Only a few pure theorists adopted a "definition-proposition-proof" style, one that is now common for theoretical analysis among all types of economists. Finally, they still saw value in debating with substantivists over which methodology best grasps the essence of the economic. In their view, the essence lay in rational choice under the constraints of resource scarcity rather than in socially embedded, nonmarketized behavior such as reciprocity.

However, a new stage of formalism was in the making even earlier than the debate, exemplified in the work of Kenneth Arrow and Gerard Debreu on the modern general equilibrium theory, published in the mid-1950s (Arrow and Debreu, 1954). This stage began to emerge in its mature form in the 1980s, as indicated by a surge of research in the fields of game theory, contract theory, and mechanism design in microeconomics. The practice of

formulating economic theories today follows this methodological paradigm, which could be summarized under the phrase "formalism in terms of complete models."

In this new formalist paradigm, incompletely specified models have been replaced by completely specified models, and in the process all the flexibility and ambiguity in the model settings have been excluded. If an economic theorist today writes a model, she must clearly specify all the parameters and details, just like a mathematician who proposes a theorem or a problem waiting to be proved or solved. Along with the move to the prevalence of complete models in formalist research, the style of exposition has also changed, in that the "definition-proposition-proof" format, the standard for professional mathematicians to organize and exhibit their results, has come to dominate the exposition of theoretical analysis. The old style of formalist writing, a loose, unsystematic mix of qualitative discussion with some mathematical symbols and formulas, has given way to a more axiomatized exposition, beginning with articulating basic definitions, then formulating propositions combining these definitions and logical predicates, and finally proving these propositions in a mathematically rigorous manner. If one opens any of the most widely used textbooks of advanced microeconomics, such as that by Mas-Colell, Whinston, and Green (1995) or Kreps (2013), one will immediately encounter exposition in the form of "definition-proposition-proof" just as one would see in a textbook for mathematics majors.

At the same time, the level of mathematics used in elaborating economics theories has reached new heights, moving from calculus and linear algebra to real analysis, which happens to be the subject of a core course for senior undergraduates or beginning graduates in departments of mathematics. Compared with the courses in calculus and linear algebra targeted at nonmath majors, courses in real analysis, a subject based on set theory, exemplify the spirit of modern mathematics and highlight the importance of axiomatization and structuralization in organizing mathematical knowledge.[7] The replacement of calculus and linear algebra with real analysis as the basic mathematical language, and a rigorous and formalized style of documenting economic theories, are just the two sides of a same coin.

Precisely because contemporary formalism is exemplified by the use of complete models and the rigorous "definition-proposition-proof" type of exposition, it is preeminently idealistic, representing or referring to cases in an unrealistic/theoretical world rather than in the real/empirical world. This is in contrast to theories formulated in the vein of the old formalism, which at least give readers the impression they are talking about the state of affairs in the real world. Furthermore, since the modeling details are completely specified, and all the implications of the model are established by a system

of definitions, propositions, and their proofs, the theoretical world established through a complete model is highly structuralized. That is to say, the state of affairs in this world are strictly prescribed by the model and are thus predictive: one can be confident that from certain postulations certain conclusions must follow, or, in other words, all the causal process and mechanisms in this hypothetical world are crystal clear. It goes without saying that this kind of structuredness, tractability, and predictability can be found nowhere in the real world.

Why Are Complete Models Useful?

How should the cognitive value of the complete models method for social scientific studies be evaluated? First of all, it must be made clear that a myth, held by many economists, indeed exists, that is, the myth of equating the implications of complete models with statements about empirical facts, or equating a hypothetical world established theoretically via complete models with the real world. This myth—built on faith that the simple logic of an idealistic object can be applied to a world of contingency, ambiguity, and complexity—leads to the mistaken inference that the model predicts with virtual certainty what will happen in the real world (Huang and Gao, 2015).

A recent study by Gilboa et al. (2014) is one of the most powerful works in mainstream economics that tackles this myth. Gilboa and his coauthors argue that proposing a formalized model amounts to establishing a "theoretical case," which is not a faithful reflection of empirical objects but rather an artificial construct that is used, as if it were an analogy, to draw inferences about the real-world issues under study. They go on to emphasize that building models and making deductions based on these models are in fact not instances of "law-based reasoning," contrary to what most economists as well as their general audience may think, but instead are instances of "case-based reasoning." Thus, analyzing a model written down by a formalist is just like analyzing a real-world case collected and summarized by a substantivist from the field. In short, what a complete model provides are not law-like predictions, but case-like and localized analogies.

Bearing in mind that complete models represent theoretical cases in a hypothetical world, why should we pay attention to such models, since they do not yield direct knowledge of empirical facts and mechanisms? Why should we not simply discard the formalist approach and embrace substantivism, which is much closer to the empirical? Are the insights generated by complete models something that substantivist methods—such as direct observation of the empirical, informal/verbal conceptualization of empirical facts, and causal process tracing in field studies—cannot provide?

The cognitive value of a complete model is to a large extent rooted in its mathematical "tractability" and "manipulability," properties derived from the complete specification of model settings, and the reduction of a rich empirical context to a highly simplified and structuralized theoretical construct, which is immanent in the complete models methodology. A model is tractable when it is possible to solve it directly, or to prove propositions on critical properties of the model. Such a model can be theoretically manipulated; that is, the parameters and/or assumptions of the model can be changed to see how different causes lead to different effects. Thus, by theoretically manipulating a complete model, a series of hypothetical experiments, which by no means are actually done in the real world, can be performed, leading to a collection of hypothetical causal relationships that can be used to make inferences about real-world causal mechanisms in the sense of analogies and/or comparisons. There is little chance that such inferences can be readily derived by observational studies of real-world cases.

Tractability is a concept commonly use in mainstream economic literature to evaluate the performance of a proposed model in pure or applied economics. To describe a model as tractable is to say that the model has explicit solutions, or if explicit solutions are hard to find, propositions on the key properties of solutions to the model can be proved, so as to make nontrivial inferences with the aid of those properties. By solving the model and exploring the properties of the solutions, we can thus understand how the postulations of the model can lead to implications, and thus derive a variety of insights complementary to those obtained via substantivist methods. In fact, comparing the implications of models with empirical facts can on some occasions lead to a better understanding of the dynamics of certain aspects of the world than by simply observing the empirical facts themselves. The latter just record things as they are and cannot suggest what might happen if things were otherwise.

Manipulability is closely connected to the concept of causality in the context of philosophy of science as well as the practice of contemporary social scientific research. One can examine whether a causal relationship between a cause C and an effect E is valid, in the view of social scientists today, by manipulating C in an appropriate way and witnessing whether E will be changed. If indeed the intervention on C brings about a change in E, then the causal relationship will be considered valid, not just a non-causal, statistical association. Yet, though it can be thought of as an empirical science, it is much more difficult in economics to empirically establish a causal relationship—a task that has been raised to a high position—than in the natural sciences. For the latter, the artificially constructed environment of the laboratory makes it easy to separate out a single structure of the object under

investigation, while keeping all other confounding mechanisms controlled. However, economics as well as other social sciences has no laboratory in which to perform controlled experiments on a large scale. Though both laboratory and field experiments since the 1980s have made a significant contribution to economic studies, especially in the field of development economics (Banerjee and Duflo, 2009), most of the topics crucial in economics, especially concerning meso- or macro-level economic facts, regularities, and patterns, are difficult, if not impossible, to investigate through experiments. Hence, a complete model can play the role of a laboratory, though a hypothetical one, where causal relationships and mechanisms can be explored. Since a model's settings are fully specified, an economist can manipulate the model's parameters and see how different implications are brought about through such interventions, just as natural scientists manipulate their laboratory settings in a controlled experiment to see what new phenomena might be produced. Yet, since they are just implications suggested in very restrictive settings and prescriptions, causal relationships and mechanisms established by manipulating a model cannot be used directly to explain or predict the hows and whys of the real world. Models can only facilitate explaining real-world cases in terms of analogies and/or comparisons and must be used in a way that recognizes the boundaries of their applicability.

The state of affairs in the world conjured up by complete models, a world that is theoretical/hypothetical and highly structuralized, is strictly prescribed by the settings and implications of such models; no other possibilities are admitted. Thus, it is much easier in this mythical world to identify mechanisms and draw causal inferences than in the much more complicated, ambiguous, and contingent real/empirical world. It is precisely through this distinction that the epistemic merit of formalism in terms of complete models becomes clear—it provides unrealistic but easy-to-identify, crystal clear mechanisms and causal relationships. These can be wielded through analogies and/or comparisons to suggest which empirical cases and/or data should be selected and investigated, and which aspects or dynamics of empirical objects are deterministic for the issue at hand and should be brought to the fore. Clearly, this merit of formalism is not something that many economists tout. They believe that the rigorous (but unrealistic) deductions based on models can be equated with general laws governing how things operate in the real world. What complete models generate, however, are not predictive general laws, but auxiliary devices of cognition, and these auxiliary devices must be complemented with other necessary methodological categories, especially substantivism, to work as a whole to make social science an area of intelligent activities of human beings, as discussed below.

The Methodological Characteristics of Complete Models: A Case Study of the Two-Sided Matching Model

Using as an example a two-sided matching microeconomics model to assess the methodological value of complete models will clarify several crucial points: first, how mathematical language is used in a complete model to fully specify modeling details; second, how finding solutions to the model and analyzing their properties can help in comprehending the nature of the model as well as the aspects of the empirical world that the model strives to represent; third, how manipulating the settings of the model can help in the exploration and identification of causal mechanisms in the empirical world; fourth, how the same model can be used to analyze different kinds of empirical phenomena by changing the details of key modeling concepts while preserving the model's general mathematical forms; and finally, under what circumstances the model can generate predictive implications for the state of affairs in the empirical world, and under what circumstances this is difficult to achieve.

To digress, the two-sided matching complete model is in fact relatively new to economics. The first complete model in the history of economics was general equilibrium theory. Economists had long been trying to find a suitable theoretical expression for the intuition that market clearing and the effective allocation of production factors are guided by the price mechanism. This is precisely what general equilibrium theory seeks to formalize. In the late eighteenth century, Adam Smith provided the first, and still, impressive expression of the idea using purely qualitative analysis around the pivotal concept of "the invisible hand." In the second half of the nineteenth century, the French economist Léon Walras made a seminal attempt to formally solve the problem of market clearing using the elementary algebraic method of solving simultaneous equations.[8] The successful formulation of general equilibrium theory was finally achieved in the 1950s by economic theorists Kenneth Arrow and Gerard Debreu. Inspired by the work of mathematician John Nash, they used a state-of-the-art mathematical tool—Kakutani's fixed point theorem in algebraic topology—to establish for the first time a rigorous theory of general equilibrium of competitive markets (Arrow and Debreu, 1954). A retrospective look into this historical trajectory will be helpful in understanding how and why economics evolved from a substantivist style of research to a formalist one, and then to the use of complete models. Since the technical details of general equilibrium theory are too complicated for the purposes of this article, we turn to a more workable example, a simplified version of the two-sided matching model, to explore the features of the method of complete models.

Complete Specification of the Model

The two-sided matching model was first proposed by American economists David Gale and Lloyd Shapley in their paper "College Admissions and the Stability of Marriage" (Gale and Shapley, 1962). The model assumes an extremely simple theoretical world, consisting of only two sets of individuals. One is a set of men, *M*; the other is a set of women, *W.* Both of the sets have n individuals, that is, the numbers of men and women are equal. A generic individual from M is denoted by *m*; from W by *w*. Here we simply call them man m and woman *w*, respectively. We can also enumerate the two sets of men and women as: $M = \{m_1, \ldots, m_n\}$ and $W = \{w_1, \ldots, w_n\}$, with the subscripts denoting the first, the second, the third, up to the *n*-th element, of the set. We assume each man has a strict preference with regard to the women, and each woman has a strict preference with regard to the men. This means that any man (woman) can make a sequential ranking of individuals from the set of women (men), and he (she) can determine his (her) most-preferred, second-preferred, third-preferred woman (man), . . ., etc. We denote the preference relationships for any m and w by the symbols $\succ$ and $\succ_w$, respectively, and denote the collection of the preference relations of all the men and women simply as $\succ$. To sum up, the theoretical world constructed in the context of the two-sided matching problem can be succinctly expressed by a triple (M , W , $\succ$). We then give the triple itself a mathematical notation, say $\mathcal{T}$, resulting in $\mathcal{T} = (M, W, \succ)$. The two-sided matching model as well as the hypothetical world it constructs can be simply referred to as $\mathcal{T}$, and all the parameters of $\mathcal{T}$ have been clearly prescribed. This is precisely the meaning of complete specification.

This example also shows that a complete model necessarily refers to a simple and highly structuralized case. Obviously, the complexity of the empirical resists complete specification—only simple objects with low-dimensional heterogeneity can be fully specified. Structuralization means that the possibilities of things are strictly regulated. Each agent, in the hypothetical world established by the model, can have only one kind of action, "choice actions," that is, to choose a preferred matching partner; how each agent acts is also constrained by the specified preference relations. It turns out that these restricted actions of agents will lead to regular patterns of matchings, as discussed below.

Finding Solutions to the Model

What might this simple but completely specified model, $\mathcal{T} = (M, W, \succ)$, tell us? From the postulations of the model, what deductions can be reached? Answering these questions reveals how economists try to "solve" complete models.

To begin with, we need to establish a suitable "solution concept" to describe the possible outcomes that the basic settings of the model could lead to, which in the two-sided matching issue naturally translates into how a couple is formed between a man from the set M and a woman from the set W. It is mathematically convenient to document how matching operates between the two sets M and W by a function, μ, which maps an element of M to an element of W; for any man m of M, the equation $\mu(m) = w$ means m matches with woman w. Obviously, if no restriction is set on the selection of the function m, we would get any type of matching that we can imagine. In this situation, though the set of all possible μ would allow abundant possibilities for how m and w match, it could not provide valuable information to specify the kind of matchings that are most useful for shedding light on actual patterns of coupling relationships in the real world. In this regard, the first contribution of Gale and Shapley (1962) was that they restricted the possibilities of μ to a smaller set consisting only of μ which they defined as "stable matchings," that is, matchings without "blocking pairs"—for any paired agents m and w in a stable matching, there exists no other agents that m and w prefer more than their current partners. If there indeed exists a man m and a woman w, both having a choice they prefer more, then m and w will form a blocking pair and make the current matching collapse. Clearly, the meaning of stable matching is closely related to that of a "stable couple," in which neither the man nor the woman has sufficient incentive, rooted in the intuitions and regularities that emerge in the empirical world, to leave the marriage. Consequently, stable matching, combining ingredients of both the formal and the empirical, is a suitable candidate for meaningful solutions to the two-sided matching model.

Once a suitable conceptualization of the solution to the model is proposed, the question become: Does the solution really exist under reasonable model settings? If stable matching does not exist under many non-trivial circumstances, its value as a concept for grasping the gist of two-sided matching models must be discounted. Discussing the existence of solutions, as well as trying to find them if they exist, is the next crucial step in the practice of formalism.

The second contribution of Gale and Shapley (1962) was that they proved the existence of stable matching under a few rather weak conditions,[9] but, more importantly, they also established a method, which they called the "deferred-acceptance algorithm," to explicitly yield solutions to a large class of matching models. By using the algorithm, one can construct for any two-sided matching problems with fixed parameters at least two stable matching solutions: one is generally the most advantageous to the women, the other to the men.

Building on Gale and Shapley's seminal work, American economist Alvin Roth found that the National Intern Matching Program (NIMP), first adopted by the American medical system to match hospitals with interns and residents in 1953, is indeed a deferred-acceptance algorithm, and that the reason the NIMP worked well was because it produced stable matchings while its predecessor did not (Roth, 1984). Roth's case study of the institutional history of the American medical labor market shows that the NIMP was gradually formed after decades of trial and error, without any clear guidance from economic theories. In sharp contrast, Gale and Shapley's algorithm was purely theoretical, involving only a mathematical deduction, and for a long time went virtually unnoticed by students of labor markets.

Manipulating the Model Settings

Having clarified the details of the model settings and learned the properties of the solutions to the model, economists can then manipulate the model settings to see how solutions change under different specifications. This is a particular advantage of the complete models method, which has become an irreplaceable methodological category in the social sciences, because it makes possible useful inferences about issues that are difficult to examine by what is considered the most reliable way to identify causal mechanisms in contemporary philosophy of science, namely experimental methods.

For instance, an important question in two-sided matching problems is which side is better off if the number of agents on a certain side increases. Questions of this sort are abundant in real life: in the matching between hospitals and residents, between landlords and tenants, between firms and workers, and between women and men, and so on, the population on one side or the other could increase for a variety of reasons. The power of matching theory lies in its ability to prove, under quite general conditions, that the side without an increase in its numbers would be in an advantageous position, while the side with an increased number of agents would generally lose out.

For another instance, Alvin Roth proposed by mathematical deduction a famous "rural hospital theorem," which shows that, in two-sided matching problems, if all the agents on one side have a low preference over some agents on the other side (with these unpopular agents vividly called "rural hospitals"), then in any stable matching these agents would not be matched (Roth, 1986). This entails an enlightening implication: stable matching will probably not lead to a good situation for agents that belong to a disadvantaged group. If one wants to make things better for them, interventions from outside would be need, since their situation would not be improved purely through the force of the matching market itself. By theoretically manipulating the model settings

and observing the effect on the model's implications, one could easily make valuable inferences on matching mechanisms, with no need to do real-world experiments, which could be costly in both time and resources, or to observe the institutional changes of real-world matching systems, which could involve a drawn-out trial and error process.

Conducting "theoretical experiments" on models by trying different specifications and applying mathematical deduction is also a natural way to expand the literature. Usually, the trajectory of the evolution of a branch of models begins with the first generation of studies, such as Gale and Shapley's groundbreaking work, shaping the initial form of the model, proposing concepts appropriate for arriving at solutions, and carrying out basic research on the properties of the solutions to the model. From that point, later studies reformulate the specifications, try different solution concepts, and explore novel implications of the model under different settings. As this practice continues, the results of a category of similar models accumulate and finally give rise to a rich literature.

Changing the Content while Preserving the Form

Once a complete model is built, the meanings that its concepts refer to can be changed while its mathematical form and all the deductions that form has generated are preserved. This exemplifies another characteristic of the complete models method, which is quite unfamiliar to scholars working in the substantivist camp. For example, in the two-sided matching model discussed above, the two sets M and W, respectively representing the two sides participating in the matching, could be given different meanings. For example, rather than men and women, they be firms and workers, hospitals and residents, landlords and tenants, organ recipients and donors, auctioneers and bidders, house sellers and buyers, and so on. Yet the nature of the solutions and the procedures to obtain them will remain largely the same.

By changing the content of concepts, a complete model can be helpful in quickly finding commonalities among numerous heterogeneous facts in the empirical world, and conclusions derived from one particular issue can be transferred to other fields to make useful inferences there. For instance, one important implication of the two-sided matching model under the setting of coupling men and women as mentioned above is that increasing the number of women is disadvantageous to the whole population of women in the marriage market. This proposition, with its mathematical proof, could also be used to infer that, for example, if there is a sudden increase in the number of tenants in the landlord-tenant matching market, they would probably also end up worse off.

Using Complete Models to Generate Predictions about Empirical Cases

We have emphasized that, generally speaking, one cannot make predictions about the real world based on complete models. But there are exceptions. When a particular case is itself highly structuralized, suitable complete models can be established to draw predictive inferences. That is to say, the possibilities in that case would be so constrained that they could be captured by a finite number of parameters, thus permitting a complete specification.

To explicate this argument, one can consider how a predictive statement could be made according to the experience of the natural sciences, the intellectual realm in which human beings have had the greatest success in inferring predictions. As a rule, two conditions need to be satisfied. First, the setting of the model should be close to the empirical case it aims to explain or reason on; and second, the factors the model does not cover should be negligible in terms of the issue under consideration.[10] In this situation, what one can deduce theoretically from the model can be understood as approximately what would occur in the empirical case.

Students of modern natural sciences have developed two approaches to satisfy these dual conditions. With respect to empirical studies, scientists since the Renaissance have relied on laboratory experiments to study natural phenomena, rather than on directly observing nature as it is. The empirical objects under examination are artificial constructions, not natural objects. The advantage of the laboratory environment is that it can greatly reduce the complexity of phenomena as they are directly given by nature and captured by observation. In the laboratory, scientists can separate out a tiny structure of the empirical, which might possibly be governed just by a single mechanism. Second, with respect to theoretical studies, scientists can develop a model that is targeted just at this pure structure of the empirical, making the two consistent in large part. The order of the two approaches, furthermore, can be reversed; that is, a model or a theory can first be proposed, and experiments can be organized around a critical implication of the model to test whether it is valid or not.

However, the way the natural sciences relate the theoretical to the empirical is rarely available to the social sciences. Few of the empirical objects of interest in the social sciences can be reconstructed in a laboratory environment. In the social sciences it is impossible to isolate and abstract a collection of basic empirical rules or laws via controlled experiments. By the same token, one can rarely build a system of theories based on empirical regularities to cover and explain a wide range of social phenomena. Even though in some branches of economics, field or laboratory experiments are applied to

investigate various objects, these experiments are most valuable as tools for examining the significance of policy interventions or the saliency of certain behavioral modes, not for drawing up a system of predictive laws.

Since social scientists cannot easily build simple, structuralized objects, they must turn to objects that are naturally simple and structuralized—that is, empirical cases with behaviors and actions that have restricted possibilities and can only operate in specified structures. In this regard, a complex system, which entails many emergentist properties, nonlinear relations, and chaotic phenomena, is certainly not a good candidate. Only in simple and structuralized empirical cases are economists able to perform their formalist techniques to build models suitable for understanding their essential structures and make approximate predictions about how things operate.

Returning to the issue of two-sided matching theory, though it is widely applied across a variety of subdisciplines of economics, on only a very few topics does it have predictive utility when it comes to the empirical. Notable exceptions are a few instances of matching: residents-hospitals, school admissions, and organ donations. In these fields matching models have reached such predictive power that they can even be used to guide the design of "markets" that do not currently exist, with the confidence that, under the settings required by the models, the designed markets for matching will generate the desired outcomes. But in some other cases, such as the marriage market and the labor market, far from providing solid predictions, matching models can only yield heuristic insights. This is because these cases are too complicated, with too many possibilities, and cannot be reduced to a few tractable dimensions. Roth and Sotomayor (1990: 2) observed that matching theory's greatest power lies in addressing themes where economics and operations research overlap. The above-enumerated issues of assignment, admission, and donation are economic themes that are among the most similar to issues in operations research.

A Reconsideration of the Substantivism/Formalism Distinction

An Extended Conception of Substantivism

Despite the prevalence of formalism in the social sciences and its epistemic merits, substantivism remains a tenacious tradition in social science methodology. However, its application in contemporary social sciences, especially in economics, is far narrower than that of formalism. One of the key reasons for the unequal influence of the two methodologies is that, compared with formalism since the 1980s, substantivism has been too emphatic in arguing

about what is *the* most essential connotation of economic behavior. Guided by this principle, substantivism has highlighted the importance of nonmarket institutions for exchange and the particularity of modes of production and redistribution in precapitalist societies and has exerted its influence in disciplines such as anthropology and economic history. But the other side of the coin is, under this self-conception, substantivism has also missed many opportunities in more general contexts of social scientific research, since most of the focus of research is on issues in contemporary market economies. In contrast, formalism has long abandoned its insistence on arguing for a certain "ontological," true conceptualization of the nature of economic behavior and has reshaped itself as purely a *method* of using the theoretical tools of complete models to investigate the empirical world. Partly because of this, it is formalism that has been reaching out from its traditional territory into that of substantivism, not the reverse.

In view of this situation, we would suggest that substantivism take a similar turn from the "ontological" to the "epistemological." Substantivism should have a broad research agenda that keeps a close eye on the empirical world and focuses more on revealing the nature of and mechanisms behind empirical objects rather than on developing idealistic/hypothetical models. We are not calling for the complete replacement of formalism with a qualitative, verbal style of research, but arguing that substantivism is a necessary complement to formalism and could serve as an irreplaceable intermediary between formal models and the empirical world—without this intermediary, social science is impossible. This reconstruction of the connotations of substantivism would amount to an extension, not a restriction, since the original focus of substantivism on nonmarket institutions and precapitalist economies—among the most suitable themes for a productive research agenda—would be preserved. A great many non-formalist practices in social studies scattered among a variety of disciplines could thus be gathered into a unified category for describing their methodological commonality.

In this extended conception, substantivism also seeks conceptualization and builds a reservoir of interrelated concepts with a certain degree of abstraction to formulate theories, instead of understanding the empirical/real world through primitive intuitions. But the substantivist approach to social studies is significantly different from the formalist approach: substantivist theories are mainly written in natural language, rather than formal, mathematical language, and descriptions in substantivist theories allow for ambiguity and contingency, in contrast to the complete specification of all the details, settings, and parameters in formalist models. In other words, substantivist theories are less structuralized than formalist theories, and thus the state of affairs they describe and discuss allows more possibilities than do formalist theories,

with their high degree of structuredness of the hypothetical world. Precisely because of this, theories formed in a substantivist way are much closer to the empirical world, a more complex place than any theory can capture. In some outstanding substantivist works, it is even difficult to distinguish theoretical analysis from empirical discussion since the two are so seamlessly knitted together. In contrast, in formalist works, not only are the theoretical and the empirical contents of research clearly distinguished from each other, but even in the writing of scientific articles the "model" part and the "data/empirical" part are sharply separated.

In order to achieve the greatest consistency and also unified, exclusive paradigms, formalism strives to develop theoretical systems hierarchically from basic concepts and axioms up to high-level propositions. Substantivism, however, usually draws ideas, insights, and viewpoints from different theoretical camps and traditions, suspending the resolution of any possible incongruities among heterogeneous theories and analytical frameworks, and thus is a more inclusive and less "disciplinary-imperialist" style of academic practice. Substantivist scholars can use in the same study concepts from Marxism, neoclassical economics, and various other sociological and anthropological camps without incurring a methodological burden. In fact, combining heterogeneous concepts from a variety of theories helps to boost rather than undermine the persuasive power of a work. It is hard to imagine this kind of practice in a work that adheres to the formalist tradition. Thus, compared to formalism, substantivism has richer intellectual resources to borrow from. It therefore has a keener sense of the real world.

This extended conception of substantivism is more inclusive than the substantivism of Karl Polanyi and his followers. For Polanyi, what mainly separates substantivism from formalism is that the former provides a different perspective on how empirical economic activities should be conceptualized. Polanyi argued that economic activities must be considered a unified entity embedded in the overall social system, and thus subject to the pivotal influence of culture, tradition, and custom. This position is in a sense the diametric opposite of the view that rational choice is the single most important mechanism behind economic behavior.[11] This substantivist notion has long been confined to the realm of anthropology and economic history, and cannot be extended into a general concept of methodology useful in a larger sphere of social scientific studies. Our redefinition and extension of substantivism, which shift the focus from the "true" nature of economic activities to a methodological function of conceptualizing complex social systems in resistance to formalization and structuralization, and intermediating between formalized models and the empirical world, instead could serve as a starting point

for reevaluating the importance of substantivism and for bridging the gap between substantivism and other methodological traditions in contemporary social sciences.

A Substantivization of Formalist Economics?

Our analysis of formalism and its recent development in the form of complete models reveals that theories formulated via the formalist approach directly describe and discuss not the state of affairs in the real world, but theoretical cases in a hypothetical world. Constructed theoretical cases are not readily equivalent to empirical/real cases and can only serve as a first step in research. The second step is to draw inspiration from the implications of complete models and to use such models to make analogies, comparisons, and inferences—all this in order to explain phenomena in the real world. In this second step, formalism as well as complete models cannot play the leading role—their epistemic function is by and large confined to the first step of research. Only substantivism can fulfill the methodological tasks in the second step.

For a concrete example of how substantivism functions as an intermediary between formalized theories and the empirical world, we return to Alvin Roth's study of the U.S. medical labor market for interns and residents. Though Roth's major contribution—a highly formalized model of two-sided matching that addresses a real-world problem—has not only been recognized but has also been hailed as an outstanding achievement in the field of microeconomics, a closer reading of his paper shows that, without a careful case study of the history of the labor market for interns and residents in the United States and a detailed examination of how interns and residents on the one hand and hospitals on the other were matched from before World War II to the 1980s, his complete model of two-sided matching could not have provided an explanation of real economic activities, and hence would not have gained such an esteemed position in the literature on the medical labor market. Roth's case study exemplifies the utility of the substantivist method in combination with a formal model in revealing the actual process of matching. It is this combination of formalism and substantivism, rather than the use of any single one alone, that played the key role in identifying the hows and whys of the successes and crises of the intern/residents–hospitals matching system in the United States.

Furthermore, it is clear that the power of Roth's two-sided matching model to explain the performance of the assignment of interns/residents to hospitals is based on the fact that the phenomenon or object the model targeted is

empirically delimited. In other words, complete models in the formalist tradition, though written in the language of mathematics and thus appearing to be highly abstract and generalized, do not seek to express themselves as a collection of general laws that describe the operation of economic systems. Rather, they reflect a constricted understanding of the boundaries for application and have empirically delimited utility. Still, in the example of the two-sided model, the implications of the model cannot be interpreted as universal laws governing all kinds of labor markets, but as applicable only in some specific contexts, such as the labor relationships between hospitals and interns/residents in certain historical periods in the United States. The high degree of mathematization and the complete specification of modeling details of the two-sided model are aimed not at presenting an ideological or truth-like metaphor independent of the empirical world, but at getting a clear and logically consistent solution to specific matching problems, as discussed in the preceding section.

This trend of using complete models of empirically delimited utility to attack micro-level, policy-oriented issues is also reflected in a shift in the center of gravity in economics in the academy. In mainstream economics since the 1980s, the area that has developed the fastest and attracted a huge number of scholars and students consists of microeconomic theory, microeconometrics, and a combination of the two. Research in this vein typically starts by using substantivist methods; draws initial empirical judgments from real cases, current or historical; then turns to formalized models to gain new insights after substantivist methods cannot proceed; and, finally, applies the implications inferred from theoretical models to the empirical, comparing model-based knowledge with reality by means of microeconometrics. Applications of this approach mainly concentrate on explaining micro-mechanisms and evaluating micro-policies, rather than describing how the whole economic system works. We can tentatively summarize this trend as a "substantivization of formalist economics" and argue that it is worth more attention as an indicating methodological movement. The history of how this trend originated and developed, and a more detailed analysis of its connotations and epistemic practice, await further study.

Conclusion

This article examines one of the most significant methodological developments in formalist economics since the 1960s formalism-substantivism debate: the use of complete models as a tool to construct and express economic theories. Though they are characterized by a high degree of mathematization, complete models are not disguised scientism that propounds idealistic universal laws, but rather are auxiliary devices of cognition with

delimited empirical utility, wielded to shed light on targeted aspects of the real world, when qualitative, verbal analysis cannot proceed. The unique epistemic merit of complete models lies in their mathematical tractability and manipulability, which can facilitate deriving hypothetical causal mechanisms to be compared with observations on processes and dynamics in the empirical world.

In an age when formalism dominates the social sciences, the importance of substantivism should be encouraged, for the two are complementary. Complete models are advantageous for their consistency in providing a basis for deducing implications from postulates, but they alone cannot link the implications to the empirical nor provide insights from this theoretical-empirical interaction. Since they are by nature simplified and structuralized, it is also inappropriate to use them to represent and analyze complex social systems that cannot be readily reduced to simple structures. Substantivism can in turn fill these methodological gaps, in that it can act as an intermediary between formal models and the empirical world and provide concepts and inferences for complex objects beyond the scope of complete models. Formalism in economics since the 1980s has in fact shown signs of a closer connection with substantivism in the subfields of microeconomic theory and microeconometrics. This epistemic trend can be called a "substantivization of formalist economics," drawing more attention from academia to investigating the relationship between formalism and substantivism today.

Acknowledgments

The author is indebted to Philip C. C. Huang, Pei Xiaolin, Chen Chuanbo, Chen Feng, and Luo Yu for their helpful criticism and suggestions.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work is supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (Grant No. 18XNB020).

Notes

1. For a recent summary of the debate as well as Polanyi's criticism, see Stanfield, Carroll, and Wrenn, 2006.
2. On the formalism/substantivism distinction in area studies, see Huang, 1985 and 1990.

3. This contrasts with the fact that, in the 1960s and 1970s, the preeminent economic journal in the English-speaking world, the *American Economic Review*, was still publishing articles from scholars using purely substantivist methods, such as Geertz (1978).
4. The distinction we draw here between complete and incomplete specification is inspired by an article by Weyl, 2019, as well as the 2015 and 2017 working paper versions of the article. In fact, in his 2015 working paper Weyl coined the term "complete models," which he offered in contrast to "price theory," the other approach to forming economic theories. But he has abandoned the term in his published article (Weyl, 2019), and has redefined it as "reductionism." In our view, since any theory will unavoidably involve reduction of the real world, which is so complex that it lies beyond the scope of any epistemic treatment, the concept of complete models is more apt at capturing the essence of the practice of theorization in contemporary formalist economics.
5. Cohen, 1994, provides a detailed investigation of the rise of the modern paradigm of natural science.
6. Morgan, 2012, contains a rich collection of examples of mathematizing and modeling in economic thinking from the era of physiocracy to neoclassical economics.
7. This spirit is articulated by the Burbaki school of mathematicians. On the impact of this school on economic methodology, see Giocoli, 2003.
8. For a summary of the development of general equilibrium theory before Arrow and Debreu's formalization, see Blaug, 1985: chap. 13.
9. Mainly the "completeness" and "transitivity" of preferences. Completeness means that any man (woman) can always confirm which is the one he (she) prefers, if asked to choose from any two of the women (men). Transitivity means if an agent prefers A over B and B over C, then she also prefers A over C.
10. These two conditions in fact are indicative of complete models as approximations of certain aspects of the empirical world, which is the traditional notion of how models function in scientific studies, as argued by Gilboa et al., 2018.
11. These ideas are most clearly expressed in Polanyi, 1977, and 2001[1944].

References

ARROW, KENNETH J. and GERARD DEBREU (1954) "Existence of an equilibrium for a competitive economy." Econometrica 22, 3: 265–90.

BANERJEE, ABHIJIT V. and ESTHER DUFLO (2009) "The experimental approach to development economics." Annual Rev. of Economics 1, 1: 151–78.

BLAUG, MARK (1985) Economic Theory in Retrospect. 4th ed. Cambridge: Cambridge Univ. Press.

COHEN, H. FLORIS (1994) The Scientific Revolution: A Historiographical Inquiry. Chicago: Univ. of Chicago Press.

GALE, D. and L. S. SHAPLEY (1962) "College admissions and the stability of marriage." American Mathematical Monthly 69, 1: 9–15.

GEERTZ, CLIFFORD (1978) "The bazaar economy: information and search in peasant marketing." American Economic Rev. 68, 2: 28–32.

GILBOA, ITZHAK, ANDREW POSTLEWAITE, LARRY SAMUELSON, and DAVID SCHMEIDLER (2014) "Economic models as analogies." Economic J. 124, 578: F513–33.

GILBOA, ITZHAK, ANDREW POSTLEWAITE, LARRY SAMUELSON, and DAVID SCHMEIDLER (2018) "Economics: between prediction and criticism." International Economic Rev. 59, 2: 367–90.

GIOCOLI, NICOLA (2003) Modeling Rational Agents: From Interwar Economics to Early Modern Game Theory. Cheltenham: Edward Elgar.

HUANG, PHILIP C. C. (1985) The Peasant Economy and Social Change in North China. Stanford, CA: Stanford Univ. Press.

HUANG, PHILIP C. C. (1990) The Peasant Family and Rural Development in the Yangzi Delta, 1350–1988. Stanford, CA: Stanford Univ. Press.

HUANG, PHILIP C. C. and YUAN GAO (2015) "Should social science and jurisprudence imitate natural science?" Modern China 41, 2: 131–67.

KREPS, DAVID M. (2013) Microeconomic Foundations I: Choice and Competitive Markets. Princeton, NJ: Princeton Univ. Press.

MAS-COLELL, ANDREU, MICHAEL DENNIS WHINSTON, and JERRY R. GREEN (1995) Microeconomic Theory. New York: Oxford Univ. Press.

MORGAN, MARY S. (2012) The World in the Model: How Economists Work and Think. Cambridge: Cambridge Univ. Press.

POLANYI, KARL (1977) The Livelihood of Man. Harry W. Pearson, ed. New York: Academic.

POLANYI, KARL (2001 [1944]) The Great Transformation: The Political and Economic Origins of Our Time. 2nd ed. Boston: Beacon.

ROTH, ALVIN E. (1984) "The evolution of the labor market for medical interns and residents: a case study in game theory." J. of Political Economy 92, 6: 991–1016.

ROTH, ALVIN E. (1986) "On the allocation of residents to rural hospitals: a general property of two-sided matching markets." Econometrica 54, 2: 425–27.

ROTH, ALVIN E. and MARILDA A. OLIVEIRA SOTOMAYOR (1990) Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Cambridge: Cambridge Univ. Press.

STANFIELD, JAMES R., MICHAEL C. CARROLL, and MARY V. WRENN (2006) "Karl Polanyi on the limitations of formalism in economics." Pp. 241–66 in Donald Wood (ed.), Choice in Economic Contexts. Kidlington, UK: Emerald Group.

WEYL, E. GLEN (2019) "Price theory." J. of Economic Literature 57, 2: 329–84.

Author Biography

Yuan Gao 高原 is an associate professor in the School of Agricultural Economics and Rural Development of Renmin University of China. He is currently working on a project on comparing the methodologies of mainstream formalist economics and classical political economy.