

SPECIAL REVIEW ESSAY SECTION

**Herbert Gintis: *Individuality and Entanglement.*
*The Moral and Material Bases of Social Life,***

Princeton, NJ, and Oxford, 2017: Princeton University Press, 384 pp.

Big Steps and Blind Spots: Herbert Gintis's Take-Over of Sociology Is Economic Imperialism

The history of the social sciences resembles an endless sequence of two different processes: fragmentation into different and even combatting camps vs integrative attempts to develop some kind of unity of the (social) sciences. This applies especially to the relation between economics and sociology, mostly in the form of a disregard for the other fields and attempts at a more or less hostile take-over. In *Individuality and Entanglement* Herbert Gintis makes a new suggestion that is partly based on some earlier contributions to this discussion. The key points involve several extensions of classical concepts of economic reasoning: the systematic inclusion of moral motives within the established framework of axiomatic rational choice theory (RCT); the extension of the classic concepts of Walras equilibrium for markets to a system of dynamic processes with temporary equilibria; skipping the assumption of isolated and sovereign actors and replacing it with the notion of overlapping networks of actors with the shared mental models of a common culture and communication; linking the emergence of moral motives to long-term processes of gene-cultural evolution; and, last but not least, systematic empirical testing of the core assumptions of these concepts in the new movements of behavioural economics and in experimental game theory in particular. In a large part of the book, this can be understood as a direct attack on sociology with the injunction to save all the valuable and indispensable

contributions sociology has made, its broad fundus of observations and conceptual descriptions of social processes and its so-called middle-range theories in particular for more analytical rigor and the use of formal instruments of economic analyses.

One of the singularities of Gintis's suggestion is that, unlike similar proposals in the past, in this case the author is familiar with sociology and even its details. Herbert Gintis knew Talcott Parsons personally and had several disputes with him early in his academic life. One chapter of the book is devoted to this exchange. It represents one pivotal point in his argumentation: consider what sociology has to contribute, i.e. norms, internalisation, and culture, but integrate these elements into the instruments of economic reasoning and model-building. This point alone highlights the contribution: It is once again a kind of imperialistic view of sociology, but clearly from a much better informed and sometimes even benevolent perspective, where the merits of sociology are appreciated much more than one is used to reading in comparable contributions from outside the camp of economics and RCT in general.

The proposed concept is summarised very clearly right at the beginning of the book. Individual chapters deal with the particular elements focused on different directions and emphases. They can be briefly summarised in 12 points:

1. A society is constituted as a gigantic game with rules which is played by human beings. The rules are not set by 'society' or by 'genes', but by coordinated individual actors, with some common

traits of the relatively stable *conditio humana* embedded in changing social networks.

2. The stability of the respective rules presupposes, as well as certain incentives to following them, the existence of unchangeable moral principles, developed in the course of the socio-cultural evolution of *homo sapiens* long before the emergence of any social institution.
3. In their concrete relations actors share specific mental models for situationally appropriate actions, driven by self-interest on the one hand and constrained by moral principles on the other hand.
4. Actions governed by interests as well as by emotions and habits can be considered 'rational', even if certain axioms of RCT are violated.
5. The respective moral motives are more than just further incentives in addition to self-interest alone. They rather function as the categorical imperatives of unconditionally valid virtues.
6. This is what creates the 'distributional efficiency' on which all social processes depend: the shared cognitions and attachments within networks in terms of culturally symbolised typical situations and reaction programmes.
7. Against this backdrop there are three basic motives: self-regarding, other-regarding, and universal motives. Universal motives refer to unconditional virtues or moral principles, whose unconditionality provides necessary stability in the face of the volatility of both egoistic and altruistic incentives.
8. There is always a trade-off between all these motives. That is the basic principle of all versions of RCT.
9. This allows RCT (and the game theory based on it) to be applied as the general microfoundation of all the social sciences, especially when the proposed extension by altruism, reciprocity, and universal moral principles is accepted.
10. The main reason why the other social sciences, sociology in particular, have not accepted this is that to date they have not yet developed a coherent analytical core and thus have not made any progress in being able to offer a competitive theoretical foundation.
11. However, it has now become possible to integrate the social sciences in the proposed manner. This would require the adoption of the general concepts and findings of theories of evolution, of the formation and change of equilibria of social networks as well as the recognition of some shared methodological rules.
12. Contributions from other disciplines would be essential as a source of necessary extensions and changes, especially if theory development were to stagnate: Recourse to historical insights and ethnographic descriptions on the one hand, and the further development of new analytical tools, such as the recently available techniques of Agent-Based-Modelling, to overcome the limitations of classical RCT on the other.

In the book's individual chapters the biological-evolutionary anchoring of rationality and morality, their embeddedness in networks, and the existence of shared cultural models are substantiated, proven, and modelled theoretically in detail and extended by the universal heritage of culture, communication, power, trust, property, and inequality as part of the universal *conditio humana*. As can already be summarised here, the book provides an admirably skilful proposal combining broad content, openness, and formal accuracy. It represents a big and important step towards confirming the perspective that a universal, explanatory-analytical, empirically well-founded, formally precise, transdisciplinary valid, acceptable, and increasingly successful social science with a universal methodological core is possible and fruitful, solves problems, and is progressive in

its insights. One could certainly object that important parts of a really 'general' social science are missing or not developed systematically enough. But anyone reasonably familiar with sociology and its core questions and approaches will easily find the necessary links. In the face of the otherwise usual nonchalance or ignorance of many economists when dealing with the results and insights of sociology, one cannot but value this highly.

Probably the most important dividing line between sociology and economics is the question of microfoundations. Economics still focuses on the relevance of costs and benefits, while, in contrast, sociology emphasises the need to consider normative-cultural orientations and what is called the 'definition' of the situation preceding any single act and suppressing any other aspect. It is a question of whether there even exists something like a categorical *un*conditionality of certain normative imperatives, as should be the case only of 'virtues' as opposed to 'preferences', for which a trade-off between different kinds of incentives should be impossible. Against this background, Gintis lists (especially in Chapter 5) the well-known refutations and anomalies of RCT. The result of these debates is also well-known: RCT is primarily based on certain formal axioms relating to consistency in behaviour, which leave everything else open, provided that these axiomatic conditions are met: the number and kinds of motives and the special cognitive processes like intentionality and orientations at the future.

Against this background there are three main arguments used to defend RCT: Most of the biases can be explained by unnoticed differences between the researcher's and the subject's interpretation of the situation, by the inclusion of additional situational aspects, such as default options and information costs, and, in particular, by the fact that such anomalies will only occur if the situation is unclear and the opportu-

nity costs of a wrong decision are low—the latter being known as the low-cost hypothesis. Gintis explains the effects of a prior definition of the situation, the framing effects, by referring to the 'state-dependency' of preferences and their relation to social context. Surprisingly, it is not then reported how this state-dependency of incentives and costs could be included in the functions of RCT. It is also not mentioned that framing effects occur empirically even in the case of high costs, as has already been demonstrated through the example of moral actions undertaken even at greatest risk [Kroneberg et al. 2010] and through experiments with the effects of punishment, to which Gintis himself refers [Fehr and Gächter 2000; Fehr and Gintis 2007]. This is probably the most serious inconsistency in Gintis's proposal: presuming the importance of the *un*questioned *validity* of moral virtues on the one hand and the *inevitable volatility* of action if there were a trade-off and choice rather than categorical *un*conditionality. This is the heart of the 'Utilitarian Dilemma' described by Talcott Parsons as the main criticism of all RCT approaches that consider 'preferences' only (see below; for the same argument from an economic perspective, cf. Heiner [1983]).

The core of Gintis's proposal is the concept of 'general *social* equilibrium', particularly in the reformulation of the sociological notion of social roles. According to Gintis, social roles are fields of partial equilibria for overlapping networks of actors. Roles are structurally constituted by incentives and the market constellations of supply and demand, but also socially and culturally through attitudes in terms of knowledge, internalised moral obligations, and behavioural readiness shared within the respective networks. It is basically the same idea as proposed by traditional sociological role theory, for example according to Linton. Social roles here refer structurally to positions supplied by organisations and filled by individual actors who match

the positional requirements with internalised obligations. Gintis refers to structural positions in general as 'social frames' and the corresponding attitudes relate above all to moral motives. The difference to traditional sociology is then the application of instruments from game theory and the modelling of roles as social equilibria. For this, the concepts of 'correlated equilibrium' and 'distributed efficiency' become key: Unlike in the case of the Nash equilibrium, where the equilibrium results from incentive structures alone, a correlated equilibrium by social roles represents a kind of external choreographer, inducing actors to mutually sort and coordinate (themselves?) through their actions within clusters of shared mental models. This happens through the transmission of certain cultural cues, which the actors obey even against their egoistic interests. These signals, however, activate moral motives, not by means of the compelling power of a 'dictator', but rather as a symbolic hint, from which actors can always deviate, especially if the incentives (to do so?) become too strong. In this way, culture functions as a 'common prior' of the shared cultural ideas connected with moral motives and thus becomes effective at establishing and maintaining social order in the form of correlated equilibria. The different forms of these equilibria then constitute the cores of the various social areas and subsystems: families, organisations, communities, public institutions, private associations, and so forth. The basis of everything is the internalisation of respective norms through the process of socialisation. It is about establishing connections between the internal preparedness for moral actions that developed *evolutionarily* on the one hand and what is *actually* expected in a particular situation and is *symbolically* signalled by the typical features of the given context.

These ideas remind one of a long-standing division within sociology and the perspectives and attempts that have sought

to overcome it. Simplifying somewhat, it is possible to distinguish three different approaches within sociology: the normative, the utilitarian, and the interpretative paradigm. Talcott Parsons represents the normative paradigm, James S. Coleman the utilitarian one. Both are relentlessly attacked by Herbert Gintis, and this for just opposed reasons: Parsons is accused of having not been able to translate, even formally, his initial ideas for a 'general' theory of action into his sociological system theory, while Coleman is criticised for not having considered moral motives as the necessary extension of RCT, even though at the time it was already more than clear that the central question of the social sciences could not be answered on the basis of egoistic motives and a focus on identifying Nash equilibria alone. Gintis can hardly restrain his criticism of both. The reason is quite clear. His proposal thrives on resolving the blind spots of these two approaches, and I, Herbert Gintis, *alone* and *first* combine both: the formal instruments of RCT (as Coleman did) *and* the moral motives (stressed so much by Parsons).

Initially at least, the concept certainly sounds quite convincing. Fehr and Gintis's article [2007] in the *Annual Review of Sociology* celebrates itself also as a kind of offensive triumph over the generally not yet evenly matching sociology. One could imagine that this would gain great support, including, of course, from the camp of economists with sociological interests within the flourishing field of game-theoretic experiments. Support can also be expected from the quite small community of RCT-oriented sociologists, for whom it would be rather easy to include what they are missing from Coleman's approach, by, for example, extending the set of preferences in the respective utility function—just like Gintis (and others) did in their experiments in order to explain cooperation where strict RCT would not have expected it. One can also read this book as follows:

should you become formally sophisticated the way we economists are, then we will be happy to adopt the 'moral' contributions of sociology—and everything will be fine. But because the formal parts of the new social theory are already at hand and it can be easily extended to include social motives, there should be no hesitation in following it, especially since Gintis himself demonstrates extensively in various chapters how this can be tackled formally in, mostly, very ingenious ways.

But is it really that simple? Those who read the book in detail will discover terms and concepts where, even after repeated reading, one doesn't know whether Gintis is using them as a friendly adaptation to sociological jargon or whether he means something completely different. This applies in particular to those chapters where the extensions of RCT are made and concerns terms such as culture, communication, moral virtues, mental predispositions, social epistemology, social framing, cultural signals, omniscient choreographer, and networked minds etc. They are all elements of his synthesis of a theory of general and correlated equilibria and mental models of cultural patterns shared in networks. One outstanding example is the concept of roles as a typical social position and the associated internalised social preferences as 'social frames'. If it was more than just jargon, one would appreciate more detailed information on how and by which processes social framing occurs, which, according to Gintis, is at the core of role dynamics and conformity to/deviation from respective standards. On this matter, you can find almost nothing in Gintis's book. On the contrary, in more than one place in the book, effects such as social framing through, for example, verbal expressions are considered to be meaningless accessories or just cheap talk that will lose their effect if the incentives for this are simply strong enough—just as one would expect from ordinary RCT in the Walras-tradition.

The strong criticism of Parsons becomes conclusive in this context. Even though Parsons, in his well-known concept of the Unit Act, added normative orientations to the elements of RCT (actors, ends, means, conditions), he failed, according to Gintis, to incorporate them systematically into his system theory. According to Gintis, he did this to ensure a niche for sociology in relation to all utilitarian approaches and particularly against economics. Here the fundamental differences become clear. According to Parsons, a 'normative orientation' is by no means just a social motive, otherwise it could be assigned to 'ends'. It is rather a basic '*orientation*' in a situation, the '*definition*' of what *primarily* counts in the given situation as ends, means, and (perceived) conditions. According to Parsons, this '*orientation*' is what explains the unconditionality of the norms on which the stability of all social order depends. Cultural symbols trigger mental models and reaction programmes as a kind of omnipotent choreographer overwriting *all* concurring motives. Thereby unconditional and collectively uniform reactions may result—as '*social equilibria*', yet without any '*choice*' between alternatives, which constitutes the core of each version of RCT.

One can see easily that this refers to the third sociological paradigm, the interpretative paradigm, which Gintis obviously considers at most to be a kind of interesting social philosophy. The interpretative paradigm is normally associated with cultural-hermeneutic sociology, without any further reference to the normative or even utilitarian paradigm and its formal means. This, however, becomes relevant in connection with Gintis's approach and his idea of culture as a choreographer of actors. It concerns two sides of his approach. First, clear and uninterrupted cultural signals ensure the immediate triggering of related responses. Second, however, these interruptions of triggered programmes in hitherto stable social equilibria can destabilise

even strong (normative) orientations and initiate a process of weighing the trade-off between gains and losses. Admittedly, the interpretative paradigm—just like Parsons for the normative one—hasn't contributed much to the analytical instruments required to model these processes of symbolic interactions and habitualised everyday behaviour, such as the performing of social roles. However, the convergences of cognitive social psychology, more recent contributions to the sociological theory of action, and recent approaches in the economic modelling of action indicate that this may be possible [cf., e.g., Bicchieri 2006; Tutić 2015]. This would be the next step towards an analytical-stringent, integrative, and truly 'general' social theory, one that in a decisive part of its contribution to explaining the emergence and stability of social equilibria is not partially blind. The still missing step would then be totally in line with Gintis's approach, summarised in the last chapter of his book: to open-mindedly address and willingly consider concepts and results from other disciplines that have already been addressing the given questions for a long time. One can only agree with this.

This confession would have been much more convincing if Gintis hadn't evaluated sociology, at least in some parts, rather unfairly and inappropriately and in the established style of economic imperialism. Parsons was aiming at more than just extending RCT by adding normative orientations to the list of ends/'preferences' in the Unit Act, and James S. Coleman deliberately (and by no means out of ignorance) used a narrowly defined conception of RCT for many key sociological problems and relevant applications and achieved milestones in many areas. Gintis's criticism of interpretative approaches is also out of place. In his view these approaches were at best forms of social philosophy, which are interesting perhaps to read but otherwise rather irrelevant. And if he was serious in asking

for tunnel vision to be avoided, this would require that he devote more systematic consideration to concepts in which internalised mental models and the coordinative or even enforcing function of symbolic structures is important and well-prepared for a strict analysis. This applies, for example, to the work of Max Weber, William I. Thomas, and Alfred Schütz, who combine structural-economic and cultural-interpretative aspects of social processes in their explanations, each of them respectively in his own way.

Finally, it would have been helpful and beneficial if Gintis had informed himself about the decades of development of the concept of 'explanative sociology', which has a strong orientation towards economics and RCT, but transcends their limits in several respects and has already taken many steps towards reaching an integrative, 'general' social theory—even going beyond Gintis's proposal. The most important example is Raymond Boudon, the first president of the European Academy of Sociology, who has always attached importance to combining the strict analytical instruments of RCT with the special features of *all* social processes: looking at the interaction of opportunities, incentives, and rational choice with the demands of institutions, norms, and values (that are) connected with the emotional and, under certain conditions, even unconditional choreographic effects of symbols and culture in general. In this book's references to sociology's contributions to a general social theory, Raymond Boudon is not even mentioned. Perhaps this was meant as a hidden compliment in the light of Gintis's overall judgement of most parts of sociology and the contributors to it as 'at best awkward', 'often simply bizarre', 'scarcely plausible', or 'doomed to failure' [Gintis 2017: 161ff.].

Hartmut Esser
University of Mannheim
hartmut-esser@t-online.de

References

- Bicchieri, C. 2006. *The Grammar of Society. The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.
- Fehr, E. and S. Gächter. 2000. 'Cooperation and Punishment in Public Good Experiments.' *American Economic Review* 90: 980–994.
- Fehr, E. and H. Gintis. 2007. 'Human Motivation and Social Cooperation: Experimental and Analytical Foundations.' *Annual Review of Sociology* 33: 43–64.
- Heiner, R. A. 1983. 'The Origin of Predictable Behavior.' *American Economic Review* 73: 560–595.
- Kroneberg, C., M. Yaish and V. Stocké. 2010. 'Norms and Rationality in Electoral Participation and in the Rescue of Jews in WWII: An Application of the Model of Frame Selection.' *Rationality and Society* 22: 3–36.
- Tutić, A. 2015. 'Revealed Norm Obedience.' *Social Choice and Welfare* 44: 301–318.

Decision and Game Theory as the Analytical Core of All the Behavioural Sciences

In *Individuality and Entanglement* Herbert Gintis, one of the leading and most dazzling figures in behavioural economics, engages in an extensive charm offensive towards sociologists. His strategy of persuasion is twofold. On the one hand, he puts forward an explicit metatheoretical argument, according to which sociology and other problem children in the social sciences such as anthropology and social psychology need to adopt the analytical core of behavioural economics or otherwise face the ugly truth of being scientifically immature. On the other hand, Gintis attempts to seduce the reader with concrete theoretical studies of quintessentially sociological concepts and phenomena.

The book is not so much a monograph as it is a collection of essays dealing with more or less interrelated themes. Many of the chapters employ dynamic models from evolutionary game theory to study facets of gene-culture coevolution and, in particular, the emergence of social 'facts' such

as property rights (Chapter 8) and social norms (Chapter 10). Chapter 3 stands out, since it describes a static model that provides a new solution to the (in)famous paradox of voting based on the idea of distributed effectivity, i.e. the idea that in large elections actors behave as if they were participating in a small election. In all these more applied chapters, Gintis metatheoretical argument is left implicit; apparently these studies are included in the book to demonstrate the power of Gintis's approach by way of example. Finally, there are four chapters (5–7 and 12) in which Gintis takes the bull by the horns and explicates his methodological claims.

This book is full of original and thought-provoking ideas about how to apply the framework of decision and game theory to the study of social phenomena. From the many substantive ideas—social rationality, three kinds of motivations (selfish, non-selfish, universalistic), private versus public personae, etc.—let me briefly just highlight one: modelling social norms via correlated equilibria, since I believe it is among the most profound ideas for social theory.

Let's recall that Rational-Choice Sociology has basically put forward three ideas on how to use decision and game theory to study social norms. First, internalised social norms can be modelled as arguments in a social utility function. Second, the theory of repeated games, and in particular so-called folk theorems, can be interpreted as providing conditions under which social norms emerge. For instance, if two actors play an indefinite number of rounds of a Prisoner's dilemma, there exist equilibria in which both players de facto do cooperate in each round. The strategies that are typically used to show the existence of these cooperative equilibria, such as the Trigger of Tit-for-Tat, can be interpreted as social norms involving some form of reciprocity and endogenous sanctioning [Voss 2001]. Third, signalling models provide