Five Principles for the Unification of the Behavioral Sciences*

Herbert Gintis

Santa Fe Institute and
Central European University

*Adapted from The Bounds of Reason: Game Theory and
 The Integration of the Behavioral Sciences (Princeton University
 Press 2009)

Disarray of the Behavioral Sciences

- The behavioral sciences (biology, economics, sociology, anthropology, psychology, political science) are in disarray, with incompatible models of human behavior across disciplines.
- We now have the analytical and empirical basis for beginning to construct an integrated behavioral science.

Four Incompatible Models of Human Choice and Strategic Interaction

1. Economics: Homo economicus, the self-regarding maximizer with unlimited and costless information processing capacity, who acts prosocially when the incentives align with selfish motives (invisible hand).
2. Sociology: Homo sociologicus, the prosocial actor socialized to fill social roles (the oversocialized individual).
3. Biology: The fitness maximizer whose prosociality is based on inclusive fitness (kin altruism) and self-interested reciprocity (reciprocal altruism).
4. Cognitive Psychology: The irrational and illogical decision-maker (the discipline's interpretation of Kahneman ,Tversky and coworkers).

Four Incompatible Models of Human Choice and Strategic Interaction

The evidence for the existence and content of these four models comes from

- (a) what is taught in introductory graduate textbooks in the discipline.
- (b) what can be assumed in a disciplinary journal article without comment or defense.

At least three of these four are wrong, and I will argue that they are all wrong, although all include fundamental insights that must be incorporated into a unified basic model of human choice and strategic interaction.

Five Principles for the Unification of the Behavioral Sciences

1. Theory of Gene-culture Coevolution (biology)
2. Socio-psychological Theory of Norms (sociology, cognitive psychology, social psychology)
3. Classical, Epistemic, Behavioral, and Evolutionary Game Theory (economics, biology)
4. The Rational Actor Model, or Beliefs, Preferences, and Constraints (BPC) Model (economics, decision theory, biology).
5. Complexity Theory

Gene-culture Coevolution

Individual fitness in humans depends on the structure of social life.
Because culture is limited and facilitated by human genetic
    propensities,
human cognitive, affective, and moral capacities are the product of
    an evolutionary dynamic involving the interaction of genes and
    culture.
References : Cavalli-sforza and Feldman 1982; Boyd and
    Richerson 1985; Dunbar 1993; Richerson and Boyd 2004;
    Bowles and Gintis, A Cooperative Species, 2009)
This coevolutionary process has endowed us with preferences that
    go beyond the self-regarding concerns emphasized in
    traditional economic and biological theory.

Gene-culture Coevolution

Gene-culture coevolution explains why we have a social epistemology facilitating the sharing of intentionality across minds,

as well as why we have such non-self-regarding values as a taste for cooperation, fairness, and retribution,

the capacity to empathize, and

the ability to value character virtues (e.g., honesty)

The Socio-psychological Theory of Norms

All social species have a division of labor, individuals being prepared for particular roles by nutritional and genetic differences.

Human society has a division of labor characterized by dozens of specialized roles,

appropriate behavior within which is given by social norms

and individuals are prepared as actors filling these roles

rendered capable through a process of socialization.

This insight goes back to Durkheim (1902), but was developed by Parsons, Goffman, and many others.

The socio-psychological theory of norms supplies mechanisms missing from game theory that promote coordinated behavior and select among Nash and correlated equilibria .

The Rational Actor Model

Evolutionary principles suggest that individual decision making can be modeled as optimizing a preference function subject to subjective beliefs and objective constraints.

Natural selection leads the content of preferences to reflect biological fitness

although the isomorphism between fitness and utility disappears outside the environment in which the preferences evolved.

The Rational Actor Model

Some caveats are in order.

Individuals do not consciously maximize something called utility, or anything else.

Individual choices, even if they are self-regarding (e.g., personal consumption) are not necessarily welfare-enhancing.

but preferences are ineluctably a function of an individual's current state.

Beliefs are the Achilles' heel of the BPC model, because the model treats beliefs as subjective, whereas individual beliefs are a part of a social network of interdependent beliefs.

Both beliefs and preferences are functions of the context of social interaction (the frame).

Game Theory

In the rational actor model, choices give rise to probability distributions over outcomes, the expected values of which are the payoffs to the choice from which they arose.

Game theory extends this analysis to cases where there are multiple decision makers.

In the language of game theory, players (rational actors) are endowed with strategies, and have certain information, and

for each array of choices by the players, the game specifies a distribution of payoffs to the players.

Game theory predicts the behavior of the players by assuming they maximize their preference function subject to the information they possess, their beliefs, and the constraints they face.

Evolutionary Game Theory

Evolutionary game theory provides the analytical apparatus for building a dynamic model of changing gene frequencies and the distribution of cultural forms.

Genes and culture obey similar dynamic laws, often captured by the replicator dynamic of evolutionary game theory.

The analogy is not perfect, however, so cultural dynamics must be supplemented by several structural principles in addition to the "imitation" mechanism at the heart of the replicator dynamics.

Society as Complex Adaptive System

The behavioral sciences advance not only by developing analytical and quantitative models, but by accumulating historical, descriptive and ethnographic evidence that pays heed to the detailed complexities of life in the sweeping array of wondrous forms that nature reveals to us.

Historical contingency is a primary focus for many students of sociology, anthropology, ecology, biology, politics, and even economics.

By contrast, the natural sciences have found little use for narrative along side analytical modeling.

Social Norms and Bayesian Rationality

- Social life comes from a double source, the likeness of consciences and the division of social labor.

  Emile Durkheim

- There is no such thing as society. There are individual men and women, and there are families.

  Margaret Thatcher

Economics models social interaction as a Nash equilibrium of a game played by rational decision-makers.

Sociology models social interaction as the role-playing of individuals guided by social norms.

Both approaches have an impressive body of evidence in their favor

Yet, each ignores the central insights offered by the other.

Bayesian Rationality and Social Norms

- I use epistemic game theory (Aumann 76),
- based on the modal logic of knowledge (Kripke, 1966)
- to establish an analytical basis for a unified model of social interaction
- based on the rational actor model (Bayesian rationality) and the psycho-social theory of norms.
- For an exposition of this approach, see The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences (Princeton University Press, 2009).

Rationality and Nash Equilibrium

- Epistemic game theory gives us a rigorous mechanism for asserting propositions as to what "rational actors" will and will not do.

- This is an improvement over the hand-waving and purple rhetoric that has plagued classical game theory.

Common Priors and the Psycho-Social Theory of Norms

- Epistemic game theory gives no plausible reason why priors should be common.
- Sociological theory, correctly but implicitly, takes the notion of a commonality of belief, based on a common culture, as an emergent property of human social systems.
- There is no way known to deduce the notion of "common culture" from lower level principles of cognition.
- The predisposition of human group members to hold a commonality of beliefs is a product of human gene-culture coevolution.

Common Priors and the Psycho-Social Theory of Norms

- An indication of disarray in the behavioral sciences is the fact that the internalization of norms---the process whereby a commonality of beliefs is secured---is not recognized by economic or biological theory.
- Neither economic or biological theory recognizes that social norms and social institutions can serve as correlating devices for the instantiation of correlated equilibria!
- It is not that economics and biology have some alternative correlating device---they simply ignore the problem.

The Harsanyi Doctrine

Game theory would have a mechanism for the formation of common priors if Harsanyi (1967-1968) were correct.

The Harsanyi doctrine holds that rational individuals can have divergent beliefs only if they have different information.

This argument is not plausible when the events involve the subjective beliefs of other agents.

The Failure of Methodological Individualism

- Methodological individualism, vigorously maintained in modern epistemic game theory, is thus incorrect,
- because we cannot derive social norms from strategic interaction, and
- we cannot derive common priors or common knowledge of the contents of minds from the interaction of heterogeneous agents, however rational and intelligent.

The Psycho-Social Basis of Common Beliefs

- I will outline an epistemological basis for the sharing of mental constructs across rational individual minds.
- First, there are natural occurrences, such as ``the ball is yellow,'' that are mutually accessible to members of a group, meaning that if one member knows x, then he knows that each other member knows x.
- This follows from no principle of rationality, and this type of inference is doubtless rare in other species.

The Psycho-Social Basis of Common Beliefs

• Second, there are higher-order socially defined events that we call games, which specify the type of strategic interaction appropriate to the social situation at hand.

• Games are not mutually accessible, but social conventions may specify that a mutually accessible event **F** indicates game **G**.

• We call **F** a frame, and we write $\mathbf{G} = \nearrow(\mathbf{F})$.

The Psycho-Social Basis of Common Beliefs

- We think of the relation "F indicates G to agent i" as asserting that when i knows F,
- i proceeds through a series of mental steps
- involving the consideration of known social regularities, such as norms and conventions,
- at the conclusion of which i knows G.

The Psycho-Social Basis of Common Beliefs

- Third, we assume that individuals are symmetric reasoners, in the sense that
- if x indicates G to one individual i, and
- if x is mutually accessible, then
- i knows that x indicates G to each other individual.

The Psycho-Social Basis of Common Beliefs

- Then, we can prove a theorem concerning common knowledge:
- Suppose x is a natural occurrence that is mutually accessible to a set of individuals, and suppose x indicates the game **G** and the individuals are symmetric reasoners, then **G** is common knowledge.

Normative Predisposition and Correlated Equilibrium

- We say an individual is has a normative predisposition, if he if he always chooses socially appropriate behavior (i.e., he follows the recommendation of the choreographer) when it is costless to do so.
- Theorem: Given epistemic game **G** with normatively predisposed players i=1,…,n, suppose **G** is common knowledge and **G** indicates social norm *N* for all players, who are symmetric reasoners with respect to **G**.
- Then, if appropriate behavior according to *N* is a correlated equilibrium for **G**, the players will choose to play this correlated equilibrium.