Rationally Inaccessible Rationality

John F. Padgett


Stable URL:
http://links.jstor.org/sici?sici=0094-3061%28198601%2915%3A1%3C26%3ARIR%3E2.0.CO%3B2-8

*Contemporary Sociology* is currently published by American Sociological Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/asa.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
Rationally Inaccessible Rationality


**John F. Padgett**

*University of Chicago*

This study, without the intent of the writer or perhaps the expectation of the reader, had at its heart this deep paradox and conflict of feelings in the lives of men. Free and unfree, controlling and controlled, choosing and being chosen, . . . forming purposes and being forced to change them, searching for limitations in order to make decisions, . . . hoping to dominate the earth and being dominated by the unseen—this is the story of man in society told in these pages.

Chester Barnard, *The Functions of the Executive*, “Conclusion” (296).

Economists, public choice theorists, and other utilitarians have achieved great strides by vigorously insisting that preferences and beliefs are exogenous. Widespread presumptions of selfishness notwithstanding, consistency and computational sophistication are the only psychological axioms required. Elster calls this the “thin theory” of rationality. The refusal to inquire into the genesis of preferences, however, creates a dilemma for utilitarian moral theory: “Why should individual want satisfaction be the criterion of justice and social choice when the individual wants themselves may be shaped by a process that preempts the choice? . . . For the utilitarian, there would be no welfare loss if the fox were excluded from consumption of the grapes, since he thought them sour anyway. But of course the cause of his holding them to be sour was his conviction that he would be excluded from consuming them, and then it is difficult to justify the allocation by invoking his preferences” (109).

In this wide-ranging and exceptionally provocative book, Jon Elster (a political philosopher from the universities of Oslo and Chicago) begins to lay the groundwork for a “broad theory” of rationality to supplement the thin. His ultimate goal is a set of criteria by which beliefs and preferences themselves can be judged rational. Elster correctly rejects the interpretation of this task as a search for end-state criteria, for that would make rationality coextensive with the true and the good. Rather, this is a search for process criteria, applicable to the history of belief and preference acquisition. Sociologists will recognize the overall project as an attempt to make precise Weber’s notion of substantive rationality, in a manner not unlike economists’ success in making rigorous Weber’s concept of formal rationality.

*Sour Grapes* approaches this task through indirection. Intuitively it is not difficult to understand what is meant by the broad theory: beliefs that are derived from the “available” evidence through causal reasoning “of the right sort,” and preferences that are consciously and “autonomously” accepted. Yet Elster is quite aware of the ambiguities contained in the key phrases, and so makes an important methodological choice. Instead of prematurely trying out specific positive characterizations only to demonstrate their inadequacy, Elster decides to constrain feasible solutions by itemizing a series of belief and preference acquisition processes that ought to be judged irrational under any characterization. The broad theory of rationality is to be approached through a contemplation of irrationality.

The crucial prolegomenon is a demonstration that the thin theory of rationality is necessarily incomplete, for if the broad theory is merely a corollary of the thin, the whole project is unnecessary. Elster, however, evokes Sidney Winter’s (1964) “general argument for satisficing” to point out that the thinly rational search for information to better ground one’s beliefs (through, for example, Bayesian updating) always presupposes a “higher” set of beliefs, to define the search. These beliefs of course can be evaluated in turn, but only by invoking another set. The regress implies that there always exists a set of primitive beliefs that are inaccessible to evaluation by (thinly) rational means. “Rational behavior can be characterized as optimizing only—or at most—with respect to given beliefs about the world, but the principles of rationality governing belief acquisition cannot be spelled out in terms of optimization” (18). The argument relies on the facts (1) that belief acquisition is not a one-shot choice, but rather an infinitely receding hierarchy of choices, and (2) that eventually the choice is no longer whether A or B is true, but whether A or the abyss.
Elster presents no analogous argument for preferences, perhaps because there exists no thin theory of preference updating to delimit. It is not hard, however, to conjecture one: People can always adjust their preferences to be transitive through appeal to metapreferences, but successive application of this technique across levels in no way guarantees transitivity. In an infinite tree, consistency can be insured only from the perspective of a given level. The thin theory of rationality presupposes consistent preferences but can offer no algorithm to decision makers for attaining them.

These arguments establish the inaccessibility of the thin theory’s primitives to evaluation by its own criteria, but they offer no help toward understanding how metabeliefs and metapreferences could nonetheless be judged substantively rational. It is clear to Elster that the process criteria desired are something like judgment and autonomy, but the meaning of these concepts must be approached through cataloguing instances of their violation. Crossing affective and cognitive end-states with affective and cognitive causes of distortion yields a tentative, although not exhaustive, short list of candidates: adaptive preference formation, preference change by framing (Tversky and Kahneman, 1981), wishful thinking, and inferential error. The emphasis is not on the essentialist chimera of an individual free from social influences, but rather it is on the psychological mechanisms by which social causes are transformed into beliefs and preferences. Causal relevance and consciousness are the necessary (but not sufficient) operational standards Elster uses to evaluate such mechanisms.

Elster’s most innovative contribution to social psychology lies in the area of adaptive preference formation. Preference changes due to feedback from actual choices or realized consequences are well known in the cognitive dissonance literature, but “sour grapes” is the adaptation of preferences to the feasible alternative set. Socially important examples of this adaptation of goals to possibilities easily spring to mind, but Elster is most concerned with deciding when to call such adaptation rational. Two suggestions are offered: Consciousness is the criterion by which nonconscious sour grapes can be distinguished from self-conscious character planning. “It is the difference between preferences being shaped by drives or by meta-preferences” (117). A second possibility is outcome autonomy, defined as the absence of any strong preference inversions in the global alternative set due to changes in the feasible subset. The second suggestion is more precise, but the first is more consistent with Elster’s process-based project.

Wishful thinking (the adaptation of beliefs to wants) is the other potentially distortive psychological process examined at some length by Elster. It is distinguished from the “decision to believe” by the criterion of consciousness, but the core of its irrationality is its formation of beliefs through causally irrelevant mechanisms. Irrationality in this process sense of course implies nothing either about the usefulness or truth of the beliefs so formed or about the instrumental value of the psychological process itself (e.g., self-fulfilling prophecies). Indeed, the most potentially powerful argument of Elster in this regard is that interest-based theories of ideology must have some nonrational belief formation process at their psychological core to be intellectually coherent.

The “by-product” basis of this conjecture deserves review in its own right. Ideological (that is, interest-induced) beliefs are for Elster essentially by-products: “states that can never be brought about intelligently or intentionally, because the very attempt to do so precludes the state one is trying to bring about” (43). Simple examples are “Be spontaneous” or “Forget it.” The relevant one-person case here is “decision to believe,” the archetype of which is Pascal’s wager: Given that I do not believe God exists, but even that I believe it is in my interest to believe God exists, can I believe God exists? The logical rather than empirical answer is no, unless the instrumental grounds for decision can be forgotten (itself a by-product). Belief is not a state that can be attained through will alone in full self-awareness. The analogous two-person case is the logical incoherence of believing another’s claim about belief unless one also believes the other’s belief is genuine (i.e., not ideological).

The implication of this by-product argument for interest theories of culture is that ruling classes can never successfully invent ideologies and intentionally impose them on lower classes who themselves are aware (or who can become aware) of the self-interest grounding. Vastly more plausible as an explanation of ruling class-cultural hegemony is the artificial selection mechanism, in which the ruling class intentionally inhibits the spread of independently generated beliefs among true believers. Even in this framework, however, believers believe only when they also believe that their beliefs are not grounded in ruling-class interests. Hence, a priestly caste can never serve the interests of the ruling class if
it is strictly subordinate to that class. Wishful thinking or some other nonconscious process may then coherently explain lower-class adherence to priests' genuine beliefs, but in this argument the wants that are driving lower-class adaptation are their own.

The only problem with Elster’s penetrating by-product analysis is that the conjunction of it with Winter’s argument hoists Elster’s project by its own petard. (It does at least if the goal is even to imagine a fully substantively rational person.) The necessary but not sufficient conditions for Elster’s substantive rationality are causal relevance and consciousness. Logical belief systems are hierarchies, one level at least of which is indiscernible by causal relevance considerations. The substantive rationality criterion of consciousness requires that that metabelief level, whatever it may be, be attained through a decision to believe. But since decisions to believe are by-products, logical belief systems remain at their core inaccessible to broad as well as to thin rationality. Or to put it another way, rigorous application of substantive rationality to the primitives of a belief system causes the whole system to unravel. No amount of evidence or thought can eliminate the necessity for those unjustifiable axioms that give that evidence and that thought structure.

The question remains whether the same holds for preferences. Elster believes that character self-planning is a coherent project, in part because of a consciousness distinction between drives and metapreferences. The distinction seems more tenuous the higher the preference hierarchy one ascends, but in any event the coherence of the project depends upon interpreting the command “Be autonomous” as not itself by a by-product. An archetypal test case, cited by Elster in a different context, is the decision to love: Given that I do not love someone, but given that I think I should, can I (in full self-awareness) truly love that person? Much depends on how one interprets “truly, but Elster (and I) say no, a conclusion that contradicts his belief in the efficacy of managing highest-order preferences. The many examples of successful character planning cited by Elster all rely on nonstandard causal chains (e.g., right for the wrong reasons), thereby violating substantive rationality’s causal relevance criterion. The examples demonstrate simply that there is more to intelligence than logical intentionality (cf. March, 1978).

These arguments about the ultimate inaccessibility of the cores of both broad and thin rationality to themselves strike me as counsel both for humility in rational choice theory and for rejoicing in us. Nonstandard causality and nonconsciousness are the psychological mechanisms that give us access to those by-product states that allow rationality to become well defined. One side of Elster clearly agrees.

Other Literature Cited

Rocky Road for Socialists

David Knoke
University of Minnesota

How to manage the transition to democratic socialism within liberal capitalism? This question is the essence of social democracy’s historical project over the last century. By rejecting the Leninist oppositional model in favor of a reformist strategy within parliamentary institutions, socialist parties throughout advanced industrial nations have staked their futures on a particular theory of political arithmetic: that a viable coalition could be forged between the numerically stagnating manual workers and the rising strata of white-collar and professional workers. Failure to solve this class-formation equation inevitably dooms socialist parties to ineffective minority status.

Debate about the size of the true working class (latent and potentially mobilizable) has