



Is Talk Really Cheap? Prompting Conversation Between Critical Theory and Rational Choice

Author(s): James Johnson

Source: *The American Political Science Review*, Vol. 87, No. 1 (Mar., 1993), pp. 74-86

Published by: American Political Science Association

Stable URL: <http://www.jstor.org/stable/2938957>

Accessed: 03/01/2010 15:44

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. 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/action/showPublisher?publisherCode=apsa>.

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 a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Political Science Association is collaborating with JSTOR to digitize, preserve and extend access to *The American Political Science Review*.

<http://www.jstor.org>

IS TALK REALLY CHEAP? PROMPTING CONVERSATION BETWEEN CRITICAL THEORY AND RATIONAL CHOICE

JAMES JOHNSON *University of Rochester*

Critical theory and rational choice theory share both overlapping concerns and parallel theoretical weaknesses. Specifically, both critical theorists and rational choice theorists are preoccupied with determining what rational can mean in the realm of social and political interaction. I show in a provisional way how game theory extends and deepens the critical theorists' basic intuition that unembellished strategic rationality cannot adequately sustain social and political interaction. And I suggest how critical theory identifies a mechanism underlying the force of the "cheap talk" that game theorists introduce in hopes of circumscribing the indeterminacy generated by their models. My goal is to stimulate productive conversation between what are typically considered discordant research traditions.

In fact, the first task of rational choice theory must be to circumscribe its own limits.

—Elster, *The Cement of Society*

Critical theory does not relate to established lines of research as a competitor, . . . it attempts to explain the specific limitations and relative rights of those approaches.

—Habermas, *The Theory of Communicative Action*

One seeks a midwife for his thoughts, another someone whom he can help: origin of a good conversation.

—Nietzsche, *Beyond Good and Evil*

Critical theory and rational choice theory—exemplified by Jurgen Habermas' theory of communicative action and noncooperative game theory, respectively—converge in improbable but potentially productive ways at the intersection of the three epigraphs. This convergence is improbable because critical theorists and game theorists are at best indifferent to each others' work. It is potentially productive insofar as, by demonstrating *mutual* relevance and *common* concerns, it not only charts a tentative course beyond indifference but, in the process, extends the promise of enhanced theoretical understanding of social and political interaction. I hope that calling attention to the convergent tasks of critical theory and rational choice can instigate a "good conversation" between these apparently discordant research traditions. I will consequently begin by indicating in a preliminary way why this intention is not quite so farfetched as both critical theorists and game theorists might at first suspect.

Habermas explicitly locates his theory of communicative action in the lineage of critical theory represented by Theodor Adorno, Max Horkheimer, and Herbert Marcuse.¹ He nonetheless reproaches earlier versions of critical theory on three interrelated counts:

In the first place, Critical Theory never really took the theoretical contributions of the social sciences and analytical philosophy seriously. It never engaged with them systematically, as it should have done, given its own

intentions. Hence, secondly, it took refuge in an abstract critique of instrumental reason and made only a limited contribution to the empirical analysis of the over-complex reality of our society. And finally, it failed to give an unambiguous account of its own normative foundations, its own status. (quoted in Dews 1986, 49)

I will not consider whether this indictment is fair. It is more important, for present purposes, only to note that Habermas seeks to distance his theory of communicative action from the work of his predecessors on each of these counts.

Consider Habermas' response to these complaints in reverse order. First, he advances the notion of communicative reason as the centerpiece of "a social theory concerned to validate its own critical standards" (1984, xxxix). Taking strategic rationality as a counterpoint, Habermas depicts communicative reason—embodied in validity claims to truth, rightness, and sincerity that are implicitly and necessarily raised in human speech—as an unavoidable pragmatic presupposition of language use and hence of social interaction. By thus providing "a systematic grounding of the concept of reason" he hopes to establish the normative basis for his critical theory (Dews 1986, 49). Second, Habermas insists that strategic rationality cannot by itself successfully coordinate social and political interaction (White 1988, 25). In order to avoid resorting to "abstract critique," however, he anchors an encompassing theory of action in the concept of communicative reason. In this way, he aims simultaneously to concede the importance of strategic rationality and dislodge it from the center of theoretical attention. Finally, Habermas expressly distances his theoretical enterprise from the traditional self-conceptions of philosophy. He casts it instead as a "research program" intended to contribute to an empirically oriented critical *social theory* (1984, xxix, 274; 1987b, 375). In this way, he aspires to take seriously the "theoretical contributions of the social sciences," suitably disencumbered of what he considers misplaced positivist pretensions.

Given these intentions it is somewhat surprising

that critical theorists do not engage game theory (surely the most systematic social scientific analysis of strategic action) in anything like a serious manner. Habermas concedes the value of game theory for understanding strategic action only in passing (1984, 86, 88n.; 1991, 242). Other critical theorists strike an even less concessive stance. If they consider the matter at all, they regard game theory as a quite foreign undertaking. Convinced of what, following Habermas, we might call the "specific limitations" of game theory, critical theorists fail to concede its "relative rights."

Thomas McCarthy, whose work is in other respects exemplary, is representative here. He announces, also in passing, that he will avoid "rehearsing the familiar debates concerning game theoretical approaches to the general theory of action" and dismissively depicts such approaches as among the "hoarier forms" of "modern social theory" (1991, 65-66). Like other critical theorists, he admonishes game theorists for reductively "conceptualizing social relations as strategic relations and social interaction as strategic interaction" and seems not to consider what they might learn from such an exercise (*ibid.*).

Despite such resistance, critical theorists might indeed learn from noncooperative game theory.² The latter constitutes a thoroughgoing attempt to probe the nature and limits of strategic rationality by reconstructing the performance of strategically competent actors in settings where binding communication is precluded.³ The unintended upshot of this effort is to demonstrate systematically what critical theorists suspect, namely, that in many settings strategic rationality alone does not suffice to sustain social relations. In fact, formal game-theoretic results show that in dynamic settings, strategic interaction generates widespread indeterminacy in the form of multiple equilibria and attendant coordination problems. This disquieting conclusion, especially in light of how some game theorists seek to remedy it by incorporating communication into their models, in turn provides warrant for suspecting that something like communicative reason as Habermas envisions it exists and functions to coordinate social and political interaction.

This prospect might well entice critical theorists into a conversation with game theorists. But what of game theorists? What might they learn from the theory of communicative action? The answer to these questions comes into sharper focus if we recall the problem of whether and how rational choice theory might circumscribe its own limits. The "scope of game theory itself is challenged only when a critic calls for consideration of a factor that is intrinsically impossible to represent in a game theoretic model" (Myerson 1992, 66). Leaving aside, for now, vagaries surrounding the phrase *intrinsically impossible*, the indeterminacy manifest in multiple equilibria seems to mark one important limit to game theory (Kreps 1990a, 95-102; Myerson 1992, 67-68). As just mentioned, game theorists themselves try to build communication into their models as a way of circumscrib-

ing indeterminacy (Crawford 1990). But they do not offer a compelling account of the force of communication in coordinating interaction. Critical theorists not only propose a mechanism underlying the force of communication but do so in ways that in important respects, are consonant with game theorists' enterprise.

My argument will be provisional, following a conversational pattern that focuses first on weaknesses in the theory of communicative action, then turns to limits of game theory, and finally (with these limits and weaknesses in mind) explores the terrain on which a conversation between the theory of communicative action and game theory might be joined. My intention, however, is not to simulate a conversation between critical theorists and game theorists. It is to provide reasons why they might pursue one themselves.⁴

VALIDITY CLAIMS IN THE THEORY OF COMMUNICATIVE ACTION

Habermas seeks to establish normative foundations for critical social theory in a "comprehensive concept of communicative rationality" (1984, 14). He insists that rationality is not a disembodied notion, that it "has less to do with the possession of knowledge than with how speaking and acting subjects *acquire and use knowledge*" (*ibid.*, 8; *idem* 1987a, 314). In particular, he claims that the primary criterion for ascertaining the rationality or otherwise of an action is whether it can be defended in the face of criticism (1984, 16).

Habermas anchors his entire theoretical enterprise in a categorical distinction between two sorts of purposive social action (1984, 84-101, 286-87).⁵ Strategic action, on his account, is oriented toward success. It coordinates interaction via influence. Communicative action, by contrast, operates in the medium of language and is oriented toward reaching understanding. It coordinates interaction via consent or rational agreement. Parties to communicative action aim cooperatively to negotiate shared understandings of the nature of their interactions.

In communicative action, competent speakers raise and respond to "exactly three criticizable validity claims"—truth, rightness, and sincerity—depending on whether they are taking up a relation to the objective, social, or subjective world (Habermas 1984, 99, 307-8; *idem* 1991, 314). In everyday communicative practice, they do so naively and implicitly; in argument or discourse, they do so reflectively and explicitly. Communicative action derives its force from the potential for rational agreement embodied in validity claims.⁶ More specifically, it derives force from the guarantee, necessarily extended by competent speakers, to redeem, in the event they are challenged, the validity claims raised by their utterances (Habermas 1984, 302; *idem* 1985, 170; *idem* 1990, 58-59). This guarantee represents the "telos of

mutual understanding" that Habermas claims is inherent in human communication (Dews 1986, 99).

Habermas insists that the process of raising and responding to validity claims is not a contingent aspect of language use (1991, 238). It constitutes the unavoidable *pragmatic* presupposition of communicative interaction.⁷ In this sense, competent speakers have no alternative. Any social actor who attempts consistently to deny that his or her utterances raise validity claims faces a "performative contradiction." The very denial presumes criteria of valid argumentation in terms of which it can be recognized as such.⁸

This is the barest sketch of Habermas' complex argument. From it, however, one can see that validity claims are "central" to the idea of communicative action (Habermas 1984, 10). This sketch also raises a fairly obvious question. What sort of argument does Habermas advance for the existence of validity claims? Habermas' entire project rides on his ability to answer this question persuasively. His theory is plausible only to the extent that he can demonstrate how rational agreement or consent, operating in the medium of language, coordinates social interaction (*ibid.*, 278, 298; *idem* 1985, 169–70). And agreement or consent, in turn, emerges because speakers mutually recognize the binding force of the validity claims raised in their speech acts (1985, 153; 1987b, 120).

Habermas uses speech act theory to explicate the concept of communicative action and to set it off from strategic action. With certain important qualifications, he differentiates communicative and strategic action in terms of the illocutionary aims pursued in the former and the perlocutionary effects sought through the latter (1984, 295). His argument on this score has been forcefully criticized, and I will not reiterate those complaints here.⁹ The difficulty, for present purposes, is that Habermas' analysis of speech acts is not itself an argument for validity claims.¹⁰ Instead, it presumes that validity claims exist and function to coordinate social interaction: "Validity claims . . . give the illocutionary act a rationally motivating force" (Habermas 1979, 65; *idem* 1984, 304).

Habermas, then, presumes—but does not show—that there is a telos of understanding built into the validity basis of human speech. Not surprisingly, critics ask to be convinced of this (Wood 1985). It is, however, important to note that Habermas actively engages his critics to a remarkable degree and, in the process, attempts to clarify or remedy what they take to be troublesome aspects of his work (e.g., Habermas 1991). So while there is perhaps reason to be skeptical of his theory as currently formulated, there also is good reason to suspect that he will continue to address its current weaknesses. I hope (1) by taking game theory seriously, to sharpen and extend Habermas' insight that strategic rationality cannot by itself coordinate social and political interaction, and (2) to dispell skepticism about the more constructive dimensions of Habermas' critical theory by suggesting that communication, rather than some more

standardly recognized mechanism such as social norms, functions to coordinate social and political interaction.

THE IDEA OF STRATEGIC COMPETENCE

Habermas populates his theory of communicative action with actors who are competent language users. He presumes, that is, that actors are "communicatively competent" in the sense that they have a generalized capacity not simply to produce grammatical sentences but to raise and respond to the validity claims implicit in everyday language use (1979, 26–29). This, however, is only one among several sorts of "interactive competence" that he attributes to social actors.¹¹

Game theorists, by analogy, attribute to actors a "unique human capacity for strategic behavior" (Elster 1979, 2, 9–18). They populate their models with actors who are *strategically* competent, who are capable of engaging in strategic rather than merely instrumental or parametric action.¹² Game theorists, that is, presume that social actors can (1) understand that their environment partly consists of other intentional agents and (2) recognize those others as equally rational. These generalized capacities enable social actors to proceed strategically by attempting to accommodate the anticipated actions of relevant others when formulating their own plans.

Habermas, too, presumes that social agents are strategically competent in something like the sense just sketched. He implies this in his typology of action and in his remarks on the scope of reconstructive theory (see my earlier work, Johnson 1991a, 1991b). But he perhaps most clearly relies on the notion of strategic competence in his account of the development of "moral consciousness" defined as "the ability to make use of interactive competence for *consciously* processing morally relevant conflicts of action" (Habermas 1979, 88). He has in mind, specifically, the ability of social actors discursively to examine the validity—and thereby potentially establish the legitimacy—of established norms. Yet Habermas shows how the very capacity to engage in normatively governed action that such practical discourse presupposes can emerge only in intimate relation to the development of strategic competence.

Habermas uses the theory of communicative action as a framework to reconstruct the interrelated processes of social and moral development that generate mature moral consciousness. Elaborating on work by developmental psychologists J. H. Flavell, Lawrence Kohlberg, and Robert Selman, he views this as a complex learning process in which the child's chronic inability successfully to navigate particular sorts of interaction precipitates the transition from one developmental stage to the next. Habermas portrays development as a dynamic process of "creative reorganization" in which capacities constitutive of prior

stages are taken up and transformed in the transition to subsequent stages (1990, 125).

The point at which, according to Habermas, strategic competence plays a decisive developmental role is with the transition from the pre-conventional to the conventional stage of interaction. In part, this transition consists in "the restructuring of pre-conventional competitive behavior into strategic action"; at this juncture the child becomes able to appreciate how his own actions are situated in a web of social interdependencies: "At the conventional level the characteristic innovation was the actor's ability to view himself in reciprocal relation to others as a participant in a process of action and *at the same time* to step outside and observe himself as a constituent part of interaction" (Habermas 1990, 159). Moreover, the child begins to recognize others as rational agents: "Alter stops being perceived as someone whose actions are determined by shifting needs and interests and is now perceived as a subject who intuitively follows rules of rational choice" (p. 150). With the transition from the pre-conventional to conventional stage, therefore, the child becomes strategically competent.

Habermas portrays the development of strategic competence as a necessary precondition of the child's emerging capacity to engage in normatively regulated interaction. He summarizes this point as follows:

the conventional stage of interaction is characterized by the rise of two contrasting types of interaction: strategic action and norm-governed interaction. Owing to the integration of the observer perspective into the sphere of interaction, the child learns to perceive interactions, and his own participation in them, as occurrences in an objective world. This makes possible the development of a purely success oriented type of action as an extrapolation of conflict behavior guided by self-interest. At the same time that strategic action is being acquired through practice, its opposite, non-strategic action, comes into view. Once the *perception* of social interaction is differentiated in this way, the growing child cannot avoid the necessity of also reorganizing types of nonstrategic action that had been left behind in his development, so to speak, to bring them into line with the conventional level. What this means is that a social world of norm-guided interactions open to thematization comes to be set off against the background of the lifeworld. (Habermas 1990, 140)¹³

As I understand Habermas' account of this dynamic, the capacity to engage in normatively regulated action both presupposes the emergence of strategic competence and compels further reorganization of prestrategic competitive behavior (pp. 141–52). I will not examine the details of his discussion. My point here is simply to indicate that "the transformation of interest-governed conflict behavior into strategic action" is among "the structural preconditions of communicative action" (p. 158).

TWO DIFFICULTIES

Strategic competence is, on Habermas' account, crucial to the development of full moral consciousness.

This provides critical theorists with a forceful reason to take seriously game-theoretic accounts of how strategic competence manifests itself in performance. We can fortify this reason by considering two difficulties that the idea of strategic competence reveals in the theory of communicative action.

First, Habermas' account of sociomoral development recapitulates a basic deficiency in his theory. Specifically, his categorical distinction between action oriented to success and action oriented toward reaching understanding is insufficiently nuanced. In his typology of rational action, for example, this misleads Habermas into depicting strategic action as essentially egoistic and atomistic (Johnson 1991a). In the present context, it similarly prevents him from fully appreciating the degree to which "strategic action" actually "comes to be *differentiated* from competitive behavior" in the developmental process (Habermas 1990, 150; my emphasis). This shortcoming manifests itself as a crucial ambiguity at the center of Habermas' reconstruction of sociomoral development.

Habermas insists that in developmental models such as his own, the transition from one stage to another involves thoroughgoing reorganization of the principles governing the earlier stage; and he sharply criticizes Kohlberg for failing to demonstrate how the latter's theory of moral development articulates such a stage structure (Habermas 1990, 127–28). Yet, in the long passage just cited, Habermas depicts the emergence of strategic competence as simply an "*extrapolation* of conflict behavior guided by self-interest." He then proceeds to elucidate the transition from pre-conventional competition to conventional strategic action by reference to an experimental study reported by J. H. Flavell that captures a pure-conflict, constant-sum game (*ibid.*, 148–50).

At a critical point in his treatment of moral development, then, Habermas mistakenly conflates strategic interaction with constant-sum, pure-conflict situations. Here, as elsewhere in his theory, Habermas obscures the by now commonplace observation that mixed motivations are central ingredients in most strategic interactions (Schelling 1960, 88–89). This reinforces the overly stark opposition he draws between strategic and communicative action. And because it leads him to overlook the particular ways that strategic actors presuppose a broader structure of communication in order to coordinate social and political interaction successfully, it prevents Habermas from consistently appreciating that strategic action is a form of *social* action (Johnson 1991a).

In a pure-conflict, constant-sum setting, each player is driven by self-interest to minimize the maximum payoff to the opponent, thereby ensuring himself the greatest possible return. In more routine variable-sum interactions, by contrast, players have mixed motives. The return to each depends, to some greater or lesser extent, on whether, and how successfully each is able to coordinate with relevant others.¹⁴ This need to coordinate renders variable-sum interactions truly strategic. It also establishes a

crucial role for communication in strategic interaction. The contrast with zero-sum settings is striking.

In a zero-sum game the analyst is really dealing with only a single center of consciousness, a single source of decision. True, there are two players, each with his own consciousness; but minimax strategy converts the situation into one involving two essentially unilateral decisions. . . . No social perception is involved. But in a mixed-motive game, two or more centers of consciousness are dependent on each other in an essential way. *Something has to be communicated.* (Schelling 1960, 163; emphasis mine)

Game theorists have struggled in recent years to account for how communication operates as a mechanism for coordinating social and political interaction. I claim in later sections that this provides the ground upon which critical theorists and game theorists might join in productive conversation.

Second, Habermas himself mentions an especially germane difficulty that his presumption of strategic competence underscores. He notes that—like “any approach that distinguishes competence from performance”—his reconstruction of sociomoral development raises a vexing methodological problem. In particular:

Such theoretical approaches face specific measurement problems because competence can be captured only in its concrete manifestations, that is, only in performance. Only insofar as these measurement problems have been solved can we isolate factors determining performance from theoretically postulated competences. It may be helpful to distinguish factors determining performance that, as *stimulators* or *accelerators*, must *supplement* or can accompany an acquired competence from the *braking* and *inhibiting* factors that act as filters. (Habermas 1990, 187)

The notion of strategic competence provides an instructive basis for exploring the implications of this statement. Consider two sets of experimental findings.

On the one hand, developmental psychologists have determined that as early as six to nine years of age children become able both to entertain second-order states (beliefs about beliefs, intentions about beliefs, etc.) themselves and to attribute such states to others. As the psychologists recognize, this capacity for recursive thought is an indispensable prerequisite of strategic rationality (Perner 1988; Perner and Wimmer 1985).

On the other hand, experimental evidence also suggests that even in relatively elementary negotiations, mature actors frequently do not avail themselves of their capacity for strategic thinking. Experimental subjects, that is, proceed parametrically, rather than strategically. They do not recognize others as equally rational agents whose intentions must be anticipated and accommodated as the subjects themselves formulate their own bargaining strategies (Bazerman and Carroll 1985, 252, 258–59). This sort of parametric interaction frequently generates outcomes that are undesirable from both individual and collective perspectives.

Taken together, this pair of experimental results highlights the complexities involved in identifying

the “braking and inhibiting factors” that might interfere with how strategic competence manifests itself in performance. Critical theorists would likely attribute the sorts of distorted interaction reported in the second set of experimental results to the influence of what Habermas calls “non-generalizable interests” (1984, 35). The psychologists who conducted the experiments advance a rival interpretation. They conclude, instead, that experimental subjects resort to parametric thinking in an effort to render bargaining situations cognitively tractable (Bazerman and Carroll 1985, 260).

This sort of conflict of interpretation is theoretically disconcerting. However, by appropriating game-theoretic insights, critical theorists could substantially mitigate such difficulties. To paraphrase Habermas, formal game-theoretic models capture the performance of strategically competent actors. In this sense, game theory is best understood as a reconstructive theory of strategic rationality (Johnson 1991b). Reconstructive theories, according to Habermas, provide formal analyses of the basic interactive competences of social actors (1979, 8–14; 1990, 21–42). They “acquire a *critical* function” to the extent that they “explain deviant cases” (idem 1990, 31–32). Critical theorists, for example, frequently assert that nongeneralizable interests distort or inhibit social interaction. Well-known game-theoretic results function critically by sustaining such assertions. Indeed, game theorists identify entire classes of strategic interaction that generate inefficient equilibria because relevant parties act solely out of narrow self-interest.¹⁵

Critical theory, however, cannot content itself with reconstructing “deviant cases.” The very notion of deviance implies a background of standard cases free, at least in principle, of distorting factors. In that sense, a critical reading of game theory, too, must discharge a “theoretical” function by facilitating formal analysis of standard strategic interaction (ibid., 32). At this level, however, critical theorists confront a rather peculiar situation, because formal game-theoretic models provide insight into the systematic limits of strategic rationality.

Game-theoretic results not only demonstrate that strategic competence is somehow distorted, as in the deviant cases mentioned. They also show how pervasive indeterminacy besets the strategically competent actors who populate game-theoretic models even where, as in coordination problems, the “braking and inhibiting” effects of nongeneralizable interests are minimized or bracketed entirely. This indeterminacy ensues because players face “strategic uncertainty” in the sense that they each know all of the options available to relevant others but do not know which option those others will actually pursue.¹⁶ It is this sort of uncertainty, for example, that the investigators invoke to explain why the subjects in the bargaining experiments reported earlier resort to parametric, rather than strategic, thinking. But the theoretical importance of strategic uncertainty extends well beyond those particular experimental results. It suggests that strategic competence is a pre-

carious achievement, susceptible not only to the distorting pressures of selfish motivation but also to endogenous limits, namely, the persistent indeterminacy that derives from strategic uncertainty.

Game theorists have devoted considerable energy to the task of tracing the implications of this insight in their models. They do not simply demonstrate *that* strategic rationality alone cannot adequately coordinate social and political interaction, but they demonstrate more systematically and precisely *why* this is so. In so doing, game-theoretic results extend the basic negative insight from which Habermas proceeds in a way that makes it deeper and more incisive.

INTERPRETING GAME THEORY

I shall now advance a compressed and deliberately partial history of game theory.¹⁷ I focus almost exclusively on the limits of game theory and especially on the indeterminacy that many game-theoretic models generate. A more balanced account would obviously require greater appreciation of the range of situations in which game theorists demonstrate that strategic rationality can coordinate social and political interaction (Kreps 1990a, 37–91). With that caveat in mind, game theorists will not, I hope, find my admittedly incomplete record of their enterprise entirely alien.

Game theorists, like rational choice theorists more generally, proceed from a presumption of rationality (Elster 1979, 116–17). They presume, that is, that actors are rational in the minimal sense that they are purposive and, in most situations of consequence, consistent.¹⁸ Thus, when game theorists encounter apparently irrational behavior, this methodological presumption directs them to scrutinize the broader situation or context for an indication of how the perplexing action might be construed as rational.

Situations of strategic interdependence like those captured in game-theoretic models unexpectedly complicate this methodological stricture. In strategic settings, each agent's action is contingent upon the action of relevant others. Because, for each agent, the context of interaction is constituted partly by other intentional agents, it cannot be treated as a parameter. As a result, the apparently straightforward idea of rationality no longer provides solid theoretical moorings.

Game theory is the formal study of the rational, consistent expectations that participants can have about each other's choices. It is . . . not the empirical study of how people make decisions but a deductive theory about the conditions that their decisions would have to meet in order to be considered "rational," "consistent," or "non-contradictory." Of course defining "rational," "consistent," or "noncontradictory," for interdependent decisions is itself part of the business of game theory. (Schelling 1967, 215)

Faced with this task, then, "The theory of games . . . does not assume rational behavior; rather it attempts to determine what 'rational' can mean when an

individual is confronted with the problem of optimal behavior in games and equivalent situations" (Morgenstern 1968, 62). In strategic situations, as Schelling and Morgenstern testify, the concept of rationality itself calls out for scrutiny. The objective becomes "to determine what 'rational' can mean."

Given this conception of their undertaking, game theorists rapidly encounter profound conceptual difficulties. Indeed, one critic concludes that

game theory addresses itself to the . . . problem that arises whenever an . . . actor takes into account the possible reactions to his own decisions of the other actors. To my mind, the main product of the very elegant apparatus of game theory has been to demonstrate quite clearly that it is virtually impossible to define an unambiguous criterion of rationality for this class of situations (or what amounts to the same thing, a definitive definition of the "solution" of a game). (Simon 1979, vol. 2, 486–87)¹⁹

This observation goes to the core of game theory. In order to grasp its force, we need to be clear about some elementary technical concepts.

In game theory, a "solution" consists of a definite prediction regarding what strategically rational players will do in the situation being modeled (Kreps 1990a, 29–30). Initially, it was supposed that this consisted largely of identifying an equilibrium point that would terminate the potentially infinite regress of reciprocal expectations unleashed as players sought to anticipate what other players would do and adjust their strategies accordingly (Elster 1979, 19). In game theory, an "equilibrium" identifies a strategy combination specifying each player's best response to the anticipated actions of relevant others. Once reached, an equilibrium is self-enforcing in the sense that no player has any incentive to depart unilaterally from his or her equilibrium strategy.²⁰

An equilibrium point, however, is only a necessary—not a sufficient—condition for the existence of a solution. This is because a game might admit either of no equilibrium or of several (Elster 1979, 117–23).²¹ For present purposes, consider the problem of multiple equilibria:

The difficulty lies in the fact that almost any interesting non-cooperative game—including almost any interesting non-cooperative bargaining game—will have a great many, and often infinitely many, very different equilibrium points. . . . This means that, if all we can say is that the outcome of the game will be an equilibrium point (or even that it will be a perfect equilibrium point), then we are saying little more than that almost anything can happen in the game. (Harsanyi 1977a, 102)

In short, the world captured by noncooperative game theory is plagued by an indeterminacy that, if not total, is pervasive. And game theory itself provides scant assistance in coming to terms with the indeterminacy it generates (Kreps 1990a, 97).

Perhaps the best-known example of this indeterminacy is the so-called "folk theorem" for repeated games.²² This theorem states that provided players accord sufficient weight to future interactions, any combination of strategies that ensures each player a

payoff greater than could be obtained by proceeding unilaterally can be sustained as an equilibrium. In an iterated prisoner's dilemma, for instance, it is well known that dynamic considerations can induce players to cooperate. The folk theorem suggests that in such situations repetition also fosters indeterminacy in the form of a "bewildering wealth" of equilibria on which players might potentially converge (Aumann 1981, 16).

Confronted with this sort of disquieting result, game theorists have a limited number of options:

A model with no equilibrium or with multiple equilibria is underspecified. The modeler has failed to provide a full and precise prediction of what will happen. One option is to admit that his theory is incomplete: an admission of incompleteness like the Folk Theorem . . . is a valuable negative result. . . . Another option is to renew the attack by changing the game's description or the solution concept. (Rasmusen 1989, 27)

Technical game theorists have primarily pursued the second option by devising new solution concepts.²³ They have advanced several "equilibrium refinements" intended to circumscribe the range of indeterminacy in their models.²⁴

The details of this technical work are unimportant here. What is significant is the rapid proliferation of solution concepts it has generated.

The situation is particularly confusing in respect of the noncooperative analysis of games with some dynamic structure in which the choice of one move or another during the play of the game may convey valuable information to the other players. Without pausing for breath it is easy to name at least ten rival equilibrium notions for which a serious case can be made that here is the "right" solution concept for such games. (Binmore 1990, 151)

This, as I understand it, is the current state of the art in noncooperative game theory (Kreps 1990a, 108–28). The indeterminacy revealed in the technical literature has elicited a variety of solution concepts that reflect divergent, not entirely convincing notions of "what 'rational' can mean" in strategic settings.

My intent in recounting this attenuated history of game theory is not to suggest that it is impossible *in principle* that game theorists might bring some order to the "confusing" theoretical situation in which they find themselves. Rather, I want to suggest that it dramatically underscores the importance of the *first* option Rasmusen mentions. It is plausible to treat the pervasive indeterminacy generated in game-theoretic models as what he calls "a valuable negative result."²⁵ More specifically, it is plausible to interpret this indeterminacy as identifying, however provisionally, the limits of strategic rationality as a mechanism for coordinating social and political interaction.

On this interpretation, game theorists have implicitly begun to chart the limits not only of strategic rationality but of their own theory, as well. This suggestion is not unprecedented. Some years ago, Anatol Rapoport observed, "The great philosophical value of game theory is in its power to reveal its own incompleteness. Game-theoretical analysis, if pur-

sued to its completion, perforce leads us to consider other than strategic modes of thought" (1966, 214). The point here is not that actual social and political practices generate indeterminate outcomes but that in the austere world captured by formal game-theoretic models, unembellished strategic rationality does so (Elster 1989, 85). And this, as Rapoport suggests, prompts the suspicion that nonstrategic factors account for the discrepancy between theoretical indeterminacy and more settled social and political practice.

IS TALK REALLY CHEAP?

This is the juncture where a conversation between critical theory and game theory can be joined in earnest. Game theorists might counter the chronicle I have presented by claiming that insofar as their theory at least begins to specify its own limits, they need not rely on critical theory to perform that task. They might, for instance, subscribe to the following estimate of their theoretical predicament: "All this may sound very slippery and unsatisfactory. There are no firm predictions, no falsifiability. If our theory appears not to work, we don't lose any sleep. 'Rationality is just one of the relevant factors,' we say blandly; 'here something else was at work'" (Aumann 1985, 37). Rather than endorse this impassive posture, I want to suggest in a preliminary way that critical theorists might advance a distinctive interpretation of game-theoretic results by identifying what, beyond strategic rationality, is at work in coordinating social and political interaction. The "something else" that they identify, of course, is precisely the binding force of speech acts.

In his theory of communicative action, Habermas seeks to reconstruct "institutionally unbound" speech acts, that is, utterances deriving force as mechanisms for coordinating social interaction solely from universal validity claims, rather than from the contingent normative or institutional context in which they are advanced (1979, 38–40, 60–61; 1984, 295, 440, n. 40). So, too, noncooperative game theory reconstructs "institutionally unbound" strategic action. It aspires to identify equilibrium outcomes sustained solely by the choices of strategically rational actors. For that purpose, it standardly treats institutions, culture, and so on as "incidental" or "inessential" detail and relegates them to the "'boundary conditions' that tie the modeled structure to its unmodeled environment" (Schelling 1960, 76, 106; Shubik 1982, 11–13). Game theory demonstrates how strategic interaction in so stark an environment generates rampant indeterminacy. It thus makes conspicuous the need to identify additional, nonstrategic mechanisms that coordinate social and political interaction.

Critical theorists identify just such a mechanism—the binding force of validity claims raised in communicative action.²⁶ Game theorists have indeed introduced communication between players in an effort to

circumscribe the indeterminacy generated by their models. The problem is that noncooperative game theory by definition captures a hypothetical world where even such communication as is allowed is nonbinding. This obviously imposes a severe constraint on the effectiveness of communication in game-theoretic models. It also affords critical theorists an opening from which to advance a distinctive interpretation of the limits of game theory.

The players who populate game-theoretic models have access to two analytically distinct types of communication (Farrell 1990, 3; Kreps 1990b, 388–89).²⁷ The first type consists of sometimes tacit, always costly signals that are rather easily modeled as moves in a game-theoretic framework. It is a form of what Habermas calls “linguistically mediated strategic action,” in which players concerned with their own success seek for that purpose to influence the choices of relevant others (1984, 295). Since any such message a player communicates is directly colored by his preferences over outcomes, signaling raises vexing questions of credibility that inhibit its capacity to coordinate interaction. Indeed, as noted earlier, the rival and not entirely compelling equilibrium refinements that game theorists advance are designed to circumscribe, to some extent, the chance that players will either convey or believe incredible messages (Farrell 1990, 4). This sort of communication thus serves, in the first instance, as much to accentuate, as to resolve, the theoretical difficulties that currently beset game theory.

The other type of communication available to players in game-theoretic models differs from the costly signals just discussed in two ways. First, it consists of talk that game theorists deem *cheap* on the grounds that since any information it conveys is costless and unverifiable, it has no *direct* bearing on players’ payoffs (Crawford 1990; Schelling 1960, 117). Second, it fits rather uneasily into game-theoretic models. Because cheap talk takes place during one or more rounds of *preplay* communication, it resists exhaustive formalization (Kreps 1990b, 388–89). Despite these apparent difficulties, however, experimental evidence and, more importantly, some recent formal results indicate that even in austere game-theoretic environments, cheap talk has unexpected effects (Crawford 1990).²⁸ In game theoretic models of bargaining, for example, strategic uncertainty generates multiple equilibria and attendant coordination problems.²⁹ In such settings cheap talk functions, within limits, to coordinate players’ expectations (*ibid.*).

Game theorists simply lack the conceptual resources to account for the binding force of cheap talk.³⁰ While they recognize that it seems to coordinate expectations effectively, they are at a loss to explain how it does so. In their idiom, talk is cheap precisely because it derives force from neither of the sources to which they typically attribute social and political coordination: self-enforcing equilibrium outcomes generated by individual choice or else some sort of contingent, exogenous enforcement.

Critical theory circumvents this theoretical predicament. Habermas, in effect, asks, “Is talk *really* cheap?” He answers *no*, not because communication is linked to payoffs or to exogenous sources of enforcement but because of the binding force of utterances *per se*. At this juncture, Habermas clearly pushes inquiry beyond the terrain of game theory. It is important, however, to note two things. First, game theorists themselves have escorted Habermas this far. They not only chart the limits of strategic rationality but incisively identify the nature of those limits in ways that critical theorists do not. Second, while critical theory and game theory seem to diverge sharply here, the resolution Habermas implicitly proposes to the game theorists’ predicament illuminates crucial aspects of their research agenda.

This last point becomes clear if we consider two analytically distinct factors that conspire to circumscribe the effectiveness of cheap talk in game-theoretic models: comprehensibility and credibility. Comprehensibility is, for game theorists, the analytically “more fundamental,” though the less familiar, of the two factors (Farrell 1990, 2). If players rely on cheap talk to coordinate their expectations, each must, in some minimal sense, be able to formulate messages that relevant others will understand. Consequently, when game theorists incorporate cheap talk into their analyses, they tacitly attribute “linguistic competence” to each actor in their model. They presume, that is, that the actors populating their models have mastered the grammatical rules of a shared natural language and that this mastery enables them to produce “comprehensible” sentences (Habermas 1979, 20, 26–27).

The problem for game theorists is that this assumption introduces an added source of indeterminacy into their models. Linguistic competence is a source of creativity. It consists of the capacity not simply to formulate comprehensible utterances but to do so in innovative ways and under novel circumstances. This complicates game-theoretic models insofar as it enables players to formulate *neologisms*, or unanticipated messages that are not used in equilibrium. In the context of a common natural language, neologisms can, under certain conditions, acquire a focal quality that undermines what might otherwise prove to be the equilibrium expectations of relevant players (Farrell 1990, 6–9).

Given the availability of neologisms, credibility, the second factor that constrains the effectiveness of cheap talk, confounds matters further still. Even in mildly mixed-motive interactions like those modeled as coordination problems, the divergent interests of relevant players offer an incentive to dissemble or misrepresent (Crawford 1990, 216–17; Farrell 1990, 1).³¹ Specifically, each relevant player has an incentive, however slight, to invoke an available focal neologism in hopes that doing so will destabilize extant equilibrium expectations and induce another more to their liking.

Game theorists respond to these destabilizing possibilities by specifying conditions under which equi-

libria, because they are "neologism-proof," remain viable (Farrell 1990, 2-3, 8-10).³² This response raises familiar difficulties. A neologism-proof equilibrium does not exist in every cheap talk game; and where one does exist, it need not be unique. More importantly, however, the conceptual strategy underlying this refinement is curious. Having implicitly endowed the players who populate cheap talk models with linguistic competence, this response ironically seeks to rein them in by, in a sense, stifling the creativity that their competence makes possible. Players remain capable of formulating neologisms, but the range of viable equilibria in a game consists only of those immune to their destabilizing effects.

Habermas approaches this problem in an instructively dissimilar way. The game-theoretic strategy just discussed seeks to mitigate credibility problems by circumscribing the potential for innovation inherent in linguistic competence. It imposes constraints on the effectiveness of potentially destabilizing neologisms that players might formulate. Habermas, instead, seeks to constrain the nongeneralizable interests that provide players with incentives to dissemble or misrepresent. He, too, presumes that social actors are linguistically competent. But as noted earlier, he also presumes that they are *communicatively* competent. He insists that they are capable not only of producing comprehensible utterances but of embedding their utterances, however unanticipated they may be, in a system of criticizable validity claims to truth, rightness, and sincerity (Habermas 1979, 26-27). This further competence sustains processes of contesting and redeeming validity claims that impose pragmatic constraints on the ability of players to make utterances that solely express their nongeneralizable interests.³³ It thus restricts the sorts of reasons players might provide in defense of any neologism they formulate. Here Habermas identifies what game theorists lack—a mechanism that might compellingly account for the binding force of language in strategic interaction.

On this reading, then, the theory of communicative action can account for the effects of what game theorists call cheap talk. Yet the account it offers, while suggestive, is incomplete; for if, as this reading suggests, the system of contestable validity claims that lends speech acts binding force provides a plausible mechanism for grasping the force of cheap talk, it also raises a series of pressing issues. I will mention only two. Most obviously, critical theorists need to provide a more precise idea of the conditions under which language can be expected to coordinate social interaction successfully. For example, communicative action remains susceptible to pressures of nongeneralizable interests. This is perhaps clearest in pure-conflict, zero-sum interactions where communication, including cheap talk, has no force.³⁴ But even in more auspicious circumstances, where actors need to communicate in order to coordinate their expectations, there is no guarantee, even at the reflexive level of discourse, that parties to communicative action

will reach agreement.³⁵ More fundamentally, suggesting *that* the validity basis of language works to coordinate social interaction is not the same as establishing *how* it does so. And, as noted earlier, critical theorists lack a persuasive account of the latter.

Habermas' theory of communicative action presently sheds little light on theoretical issues of this sort. This may seem to vitiate his accomplishment. But this conclusion would be, at best, misleading. Consider the relative purchase that game theory and the theory of communicative action afford us in conceptualizing the binding force of cheap talk. Each theory identifies factors that disrupt that force. But while game theory provides no account of why cheap talk can ever succeed, critical theory identifies a mechanism to explain how, within constraints, it might coordinate social and political interaction. If critical theorists have considerable work to do and have at hand only some of the conceptual resources necessary to the task, game theorists are at a loss over where and how to begin.

CONCLUSION

My arguments are intentionally preliminary. I present them in hopes of initiating a good, if unlikely, conversation between critical theorists and game theorists. Skeptics may complain that I do not report any "results." But I anticipate that a conversation of the sort I envision would be productive, as well as edifying. Where, then, do my arguments leave the respective parties to this hypothetical conversation?

A too simple reading might take my arguments as vindicating critical theory. This reading might proceed as follows. Both critical theory and game theory are preoccupied with determining "what 'rational' can mean" in the realm of social interaction. Game theorists offer a robust analysis of strategic interaction. In the process, they plot the dual limits of strategic rationality and of their own formal reconstructions of it. Habermas counsels against the sort of impassive position they seem ultimately to adopt in the face of these limits. But because, like them, he is concerned with "institutionally unbound" interaction, he resists resorting to exogenous sources, such as social norms, culture, or institutions as coordinating mechanisms. Instead, he reconstructs a comprehensive theory of rationality anchored in the concept of communicative action. The resulting theory of communicative action enables Habermas to specify the limits of game theory. It also provides the vantage point from which he aspires simultaneously to validate the normative presuppositions of his critical theory and assess the legitimacy or otherwise of prevailing social and political arrangements.

This reading is, indeed, overly simple. It suggests far greater closure than my arguments actually warrant. First, game theorists systematically explore the nature and limits of strategic rationality in ways that do not just bolster, but extend and deepen, critical

theorists' conviction that it cannot by itself sustain social and political relations. The former's results thus highlight the poverty of the latter's understanding of strategic action. Moreover, the sketch of game theory that I present is partial. It provides no reason to presume that game theorists are incapable of further clarifying the "confusing" state of their discipline and thereby broadening the range over which we might come to expect strategic rationality to coordinate social and political interaction effectively.

Second, following Habermas, critical theorists might pursue a persuasive resolution to the predicament that currently besets game theory. But this resolution surely is provisional. It might be challenged by game theorists themselves and perhaps by others who advance rival accounts of the mechanisms underlying discussion-induced cooperation. Moreover, nothing I have said here definitively establishes the strongest of Habermas' claims, namely, that there is a telos of understanding intrinsic to human language; that this telos is embodied in a system of criticizable validity claims; that these represent the unavoidable pragmatic presuppositions of human interaction; and that, therefore, in social interaction communicative reason is necessarily prior to strategic rationality. In order adequately to defend his project as he currently formulates it, Habermas must justify these claims in a more compelling way than he has to date (Johnson 1991a). My arguments simply suggest that attempting such a task is not nearly so preposterous as critics sometimes make out.

Finally, game theorists will surely be apprehensive about the conversation I propose. If overcoming impassivity demands endorsing Habermas' ambitious enterprise in its entirety, they might reasonably remain skeptical. Yet critical theory itself is not necessarily committed to his entire agenda. Even sympathetic observers suspect that "Habermas' conceptions of reason and rationalization, theory and discourse . . . are stronger than his arguments warrant or his project requires." This suspicion is surely sound. And proposals to remedy it by pursuing "an alternative 'weaker' program for critical theory" are attractive (McCarthy 1991, 3).³⁶ A constructive conversation between critical theorists and game theorists might illuminate this more modest agenda. Whether and how it might then be realized remains to be seen.

Notes

Thanks to Jack Knight, Joe Heath, Jenny Mansbridge, and J. Donald Moon for comments on earlier drafts. Jack, in particular, once again helped me see, if not entirely overcome, the shortfall from thinking one has made an argument to actually making it.

1. For a critical overview of this earlier work that connects it with Habermas' concerns, see Held 1980 and Honneth 1987. For the broader contours of Habermas' own project, see McCarthy 1978 and White 1988.

2. Game theorists distinguish between cooperative and noncooperative games. The crucial difference is that the

former allow for *binding* communication between players, while the latter do not (Binmore 1990, 32). Game theorists consider noncooperative games to be analytically fundamental (Harsanyi 1977a, 92; Rasmusen 1989, 29). Consequently, I focus my remarks on the theory of noncooperative games.

3. Philosophers frequently evaluate economic models, including game-theoretic models, solely on the criterion of empirical performance. However, such "concern with problems of empirical appraisal is exaggerated" and fails to "recognize that the activities of formulating economic models and investigating their implications are a sort of conceptual exploration" (Hausman 1989, 115). On the reading I present, game-theoretic models are precisely formal explorations of our standard concept of strategic rationality. This is an idiosyncratic rendering of the task of game theory (but see Aumann 1985).

4. A skeptic might entirely discount the prospects for the sort of conversation I propose. On such a view, theorists are so psychologically wedded to the standard explanatory mechanisms of their discipline that (under ordinary circumstances, at least) they are incapable of recognizing the plausibility of rival mechanisms (Stinchcombe 1991, 374-75). I obviously cannot promise that the conversation I propose will actually occur. But insofar as both critical theory and game theory suffer from the theoretical problems I identify, advocates of each have good reason to pursue it. Precisely because their circumstances are not "ordinary," conversation between them is possible.

5. For criticisms of this distinction as Habermas draws it, see Johnson 1991a. For some recent clarifications see Habermas 1991.

6. Habermas insists that communicative action derives its force from rational agreement or consent and *not* from either normative consensus or strategic compromise (1984, 296-97).

7. It is important to note that Habermas insists that speech acts derive force not just from *semantic* aspects of sentences but from *pragmatic* dimensions of utterances. Thus, a theory of meaning is inadequate to the task he sets himself. To this end, he advocates a formal pragmatics that can reconstruct the universal conditions of valid speech acts. This crucial point is beyond my present scope, however. See Habermas 1979, 1-68; idem 1992, 57-87.

8. See Habermas 1990, 80-81, 95-96, 116, 129-30. But Habermas explicitly recognizes that his argument from performative contradiction is not decisive (1990, 81, 95). It demonstrates that language use is not without presuppositions. However, it does not sustain any particular characterization of its presuppositions.

9. See Tugendhat 1985; White 1988, 45-46; and Wood 1985. Habermas seeks to defuse these criticisms by making certain "terminological clarifications" (1991, 239-45). These refinements do not bear directly on the present issues. It is important only to note that Habermas himself insists that his clarifications do not alter his basic distinction between communicative and strategic action (but see Bohman 1988).

10. Habermas inadvertently creates the impression that his analysis is such an argument in the following passage: "If we were not in a position to refer to the model of speech we could not even begin to analyze what it means for two subjects to come to an understanding with one another. . . . Naturally, speech and understanding are not related to one another as means to end. But we can explain the concept of reaching understanding only if we specify what it means to use sentences with communicative intent. The concepts of speech and understanding reciprocally interpret one another" (Habermas 1984, 287). To "use sentences with communicative intent," however, is precisely to extend a warranty to redeem, if necessary, the validity claims that your utterances raise.

11. For Habermas, "competences" are universal "capacities to solve particular types of empirical-analytic or moral-practical problems" (1990, 33). On the category of "interactive competence" see Habermas 1979, 73.

12. On the distinction between parametric and strategic rationality, see Elster 1979, 18-19, 117-18. Although he is inconsistent on this point, Habermas at times seems to

subscribe to something like the relatively benign view of strategic action presented here (1984, 85, 172–74). See Johnson 1991a.

13. I am indebted to Michael Neblo for bringing this passage to my attention during a conversation in May 1991 and for more generally helping to clarify my thoughts on the subject of strategic competence.

14. Game theorists sometimes acknowledge that coordination generates power (Hardin 1990, 363, 367–69; Schelling 1960, 108). Critical theorists would be concerned to question the legitimacy of resulting outcomes. Recognizing the role of communication in strategic interaction might afford them an opening from which to do so.

15. See Schelling (1978, 225–26). Inefficient equilibria are a species of collective irrationality. They are stable situations where all relevant actors proceed in such a way that all could be better off if all acted differently. The classic example is mutual defection in the one-shot prisoner's dilemma. Critical theorists, however, should note that the suboptimal outcome in a prisoner's dilemma results less from strategic rationality than from the nongeneralizable interests that subvert it. In a prisoner's dilemma, it does not matter what others will do—defection is a dominant strategy. It is in each player's interest to defect *irrespective* of the anticipated actions of others. And this, on the standard story, is because the prisoners are prevented from communicating binding commitments.

16. See Shubik 1982, 7–8 and Crawford 1990, 214, 216, 218. In game theory, a player's "strategy" must allow for every possible move on the part of other players (Rapoport 1966, 44; Shubik 1982, 34). Binmore (1990, 117–20, 176–78) and Rubinstein (1991) explore problems with this definition.

17. Habermas casts *The Theory of Communicative Action* largely as "a history of theory with systematic intent" (1984, 139–40). Toward that end, he engages in "the flexible exploration and deliberate exploitation of important theories constructed for explanatory purposes." My account in this section can be read in similar terms. For a more detailed, orthodox account see Aumann 1989.

18. See Harsanyi 1977a, 84; Schelling 1978, 17; and Myerson 1992, 66–69. This can, but need not, mean that actors are motivated by narrow self-interest (Elster 1979, 116; Shubik 1982, 81).

19. Compare Binmore 1990, 20.

20. The basic equilibrium concept in game theory is a Nash equilibrium, named for John Nash, who in 1950 proved formally that under certain conditions (specifically, that actors are allowed to play "mixed" strategies) there is at least one equilibrium in any finite *n*-person game (Shubik 1982, 242). It goes without saying that the existence of an equilibrium outcome does not entail that it is or can be normatively justified (Schelling 1978, 25–27).

21. See also Shubik 1982, 242–47. Elster (1986) discusses the implications of this point for rational choice theory more broadly.

22. Kreps is a nontechnical sketch of this result (1990b, 505–15). Fudenberg and Maskin (1986) is a technical presentation.

23. In political science, by contrast, formal theorists attempting to model elections and voting procedures have redescribed the "game" by introducing structure of various sorts into their models. See, e.g., Shepsle (1989) on the role of institutional arrangements and Elster (1989) on the effects of social norms. For a persuasive critical account of these efforts, see Knight 1992.

24. "Equilibrium refinements are strengthenings of the requirement that behavior constitute a Nash equilibrium, usually strengthenings that invoke in some manner or other the idea that players should not be allowed to make incredible threats or promises or to draw incredible inferences from things they observe" (Kreps 1990a, 108). In the passage cited earlier, Harsanyi parenthetically mentions one example—a "perfect" equilibrium—of this sort of refinement.

25. There are two reasons why one might resist this reading. The first is a more or less ideological faith that even in

areas of social and political life not governed by markets, uncoordinated individual action should generate equilibrium outcomes similar to those generated by markets. There is little reason, however, to treat unique, market-induced equilibria as more than a special case (Schelling 1978, 25–26). The second reason consists of a commitment to "positive science." On such a view, an equilibrium provides the basis for testable predictions (Harsanyi 1977b, 4; Ordeshook 1986, 98). Conversely, multiple equilibria compromise testability, thereby threatening game theory's scientific status. But this is a too narrow and, while common, not an especially compelling, rendering of game theory. See Johnson 1991b and Bohman 1991.

26. Critical theorists might resist my use of the word *mechanism* here, but this is precisely the term Habermas uses to depict the force of speech acts in coordinating social and political interaction (e.g., 1984, 101).

27. For a nontechnical survey of how communication figures in game-theoretic models of political decision making see Austen-Smith 1992.

28. The formal results are especially important for two reasons. First, experimental studies have difficulty discriminating between the possible mechanisms underlying discussion-induced cooperation (Cooper et al. 1990, 230; Orbell, Van de Kragt, and Dawes 1988). Second, the formal results suggest that speech acts per se have force as a mechanism for coordinating social interaction. They thus challenge a tacit, if common, premise of experimental studies that view discussion as important only insofar as it sustains some other factor (e.g., generalized norms, contract-based promising, social identity), which in turn coordinates interaction. See, e.g., Orbell, Van de Kragt, and Dawes 1988. While it may be plausible to see rudimentary norms or identities arising in face-to-face discussion among experimental subjects, it is difficult to imagine how they could emerge among the strategic actors who populate formal game-theoretic models.

29. Elster provides a nontechnical review of game-theoretic models of bargaining (1989, 50–96). Coordination problems are endemic to bargaining models regardless of whether they represent negotiations as a cooperative or noncooperative game (Crawford 1990).

30. This may seem to be a harsh judgement. Consider, however, a related comment by Ariel Rubinstein, who, after briefly considering recent attempts to incorporate cheap talk into game-theoretic models, concludes, "It is my impression that although language plays a crucial role in resolving conflicts, game theory has so far been unable to capture this role" (1991, 921). While this is surely true at present, we can remain agnostic about whether it is true in principle.

31. Critical theorists will not find this terribly surprising. They should recognize, however, that the distorting factor here is nongeneralizable interest and not, as they tend to presume, strategic rationality itself. This surely suggests that rather than decrying the allegedly baleful impact of strategic rationality, critical theory ought to explore how nongeneralizable interests disrupt *both* communicative and strategic interaction.

32. Recent developments more or less directly build upon this notion of neologism-proofness. Matthews, Okuno-Fujiwara, and Postlewaite (1991), for instance, explore various extensions in the context of face-to-face interactions. Myerson, by contrast, introduces institutional structure in the guise of a mediator who, by randomizing over messages, can induce players to adopt "correlated" strategies (1989; idem 1991, 249–58, 283–88). In Myerson's work social interactions are coordinated less by the force of language per se than by the institutionalized capacities of the mediator. Since I am here concerned with the force of language in "institutionally unbound" interactions, this development, while interesting in its own right, is strictly speaking beside the point.

33. Recall that according to Habermas, this process can operate implicitly in everyday communicative practice or explicitly and reflectively in discourse or argumentation.

34. In such interactions, self-interest gives each player an incentive to adopt "dramatically anticommunicative" ran-

domizing strategies precisely for the purpose "of expunging from the game all details except the mathematical structure of the payoff, and from players all communicative relations" (Schelling 1960, 105, 165).

35. The relevant actors might revert to strategic interaction and risk indeterminacy, or they might break off relations altogether (Habermas 1984, 17–18; idem 1979, 3–4).

36. For an interesting move in this direction, see Bohman 1988.

References

- Aumann, Robert. 1989. "Game Theory." In *The New Palgrave: Game Theory* ed. John Eatwell, et al. New York: W. W. Norton.
- Aumann, Robert. 1981. "Survey of Repeated Games." In *Essays in Game Theory and Mathematical Economics*, ed. Robert Aumann et al. Zurich: Bibliographisches Institut AG.
- Aumann, Robert. 1985. "What Is Game Theory Trying To Accomplish?" In *Frontiers of Economics*, ed. Kenneth Arrow and Seppo Honkapohja. New York: Blackwell.
- Austen-Smith, David. 1992. "Strategic Models of Talk in Political Decision Making." *International Political Science Review* 13:45–58.
- Bazerman, Max, and John Carroll. 1985. "Negotiator Cognition." *Research in Organizational Behavior* 9:247–88.
- Binmore, Ken. 1990. *Essays on the Foundations of Game Theory*. New York: Blackwell.
- Bohman, James. 1988. "Emancipation and Rhetoric." *Philosophy and Rhetoric* 21:185–204.
- Bohman, James. 1991. *New Philosophy of Social Science*. Cambridge: Massachusetts Institute of Technology Press.
- Cooper, Russell W., Douglas V. De Jong, Robert Forsythe, and Thomas W. Ross. 1990. "Selection Criteria in Coordination Games: Some Experimental Results." *American Economic Review* 80:218–33.
- Crawford, Vincent. 1990. "Explicit Communication and Bargaining Outcomes." *American Economic Review* 80:213–19.
- Dews, Peter, ed. 1986. *Habermas: Autonomy and Solidarity*. London: Verso.
- Elster, Jon. 1979. *Ulysses and the Sirens*. Cambridge: Cambridge University Press.
- Elster, Jon. 1986. "The Nature and Scope of Rational Choice Explanation." In *Essays on Actions and Events*, ed. Ernest LePore and Brian McLaughlin. New York: Blackwell.
- Elster, Jon. 1989. *The Cement of Society*. Cambridge: Cambridge University Press.
- Farrell, Joseph. 1990. "Meaning and Credibility in Cheap Talk Games." In *Mathematical Models in Economics*, ed. M. Dempster. New York: Oxford University Press.
- Fudenberg, Drew, and Eric Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information." *Econometrica* 54:533–54.
- Habermas, Jurgen. 1979. *Communication and the Evolution of Society*. Boston: Beacon.
- Habermas, Jurgen. 1984. *The Theory of Communicative Action*. Volume 1. Boston: Beacon Press.
- Habermas, Jurgen. 1985. "Remarks on the Theory of Communicative Action." In *Social Action*, ed. Gottfried Seebass and Raimo Tuomela. Amsterdam: Reidel.
- Habermas, Jurgen. 1987a. *The Philosophical Discourse of Modernity*. Cambridge: Massachusetts Institute of Technology Press.
- Habermas, Jurgen. 1987b. *The Theory of Communicative Action*. 2 vols. Boston: Beacon.
- Habermas, Jurgen. 1990. *Moral Consciousness and Communicative Action*. Cambridge: Massachusetts Institute of Technology Press.
- Habermas, Jurgen. 1991. "A Reply." In *Communicative Action*, ed. Axel Honneth and Hans Joas. Cambridge: Massachusetts Institute of Technology Press.
- Habermas, Jurgen. 1992. *Postmetaphysical Thinking*. Cambridge: Massachusetts Institute of Technology Press.
- Hardin, Russell. 1990. "The Social Evolution of Cooperation." In *The Limits of Rationality*, ed. Karen Cook and Margaret Levi. Chicago: University of Chicago Press.
- Harsanyi, John. 1977a. "Advances in the Understanding of Rational Behavior." In *Rational Choice*, ed. Jon Elster. New York: New York University Press, 1986.
- Harsanyi, John. 1977b. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge University Press.
- Hausman, Daniel. 1989. "Economic Methodology in a Nutshell." *Journal of Economic Perspectives* 3:115–27.
- Held, David. 1980. *Introduction to Critical Theory*. Berkeley: University of California Press.
- Honneth, Axel. 1987. "Critical Theory." In *Social Theory Today*, ed. Anthony Giddens and Jonathan Turner. Stanford: Stanford University Press.
- Johnson, James. 1991a. "Habermas on Strategic and Communicative Action." *Political Theory* 19:181–201.
- Johnson, James. 1991b. "Rational Choice as a Reconstructive Theory." In *The Economic Approach to Politics*, ed. Kristin Monroe. New York: Harper Collins.
- Knight, Jack. 1992. *Institutions and Social Conflict*. Cambridge: Cambridge University Press. Forthcoming.
- Kreps, David. 1990a. *Game Theory and Economic Modelling*. Oxford: Oxford University Press.
- Kreps, David. 1990b. *A Course in Microeconomic Theory*. Princeton: Princeton University Press.
- McCarthy, Thomas. 1978. *The Critical Theory of Jurgen Habermas*. Cambridge: Massachusetts Institute of Technology Press.
- McCarthy, Thomas. 1991. *Ideals and Illusions*. Cambridge: Massachusetts Institute of Technology Press.
- Matthews, Steven, Masahiro Okuno-Fujiwara, and Andrew Postlewaite. 1991. "Refining Cheap-Talk Equilibria." *Journal of Economic Theory* 55:247–73.
- Morgenstern, Oscar. 1968. "Game Theory: Theoretical Aspects." In *International Encyclopedia of the Social Sciences*, vol. 6. New York: Macmillan.
- Myerson, Roger. 1989. "Credible Negotiation Statements and Coherent Plans." *Journal of Economic Theory* 48:264–303.
- Myerson, Roger. 1991. *Game Theory*. Cambridge: Harvard University Press.
- Myerson, Roger. 1992. "On the Value of Game Theory in Social Science." *Rationality and Society* 4:62–73.
- Neblo, Michael. Conversation with author, May 1991.
- Orbell, John, Alphons Van de Kragt, and Robyn Dawes. 1988. "Explaining Discussion Induced Cooperation." *Journal of Personality and Social Psychology* 54:811–19.
- Ordeshook, Peter. 1986. *Game Theory and Political Theory*. Cambridge: Cambridge University Press.
- Perner, Josef. 1988. "Higher Order Beliefs and Intentions in Children's Understanding of Social Interaction." In *Developing Theories of Mind*, ed. Janet Astington, Paul Harris, and David Olson. Cambridge: Cambridge University Press.
- Perner, Josef, and Heinz Wimmer. 1985. "'John Thinks That Mary Thinks That . . .': Attribution of Second Order Beliefs by Five to Ten Year Old Children." *Journal of Experimental Child Psychology* 39:437–71.
- Rapoport, Anatol. 1966. *Two-Person Game Theory*. Ann Arbor: University of Michigan Press.
- Rasmusen, Eric. 1989. *Games and Information*. New York: Blackwell.
- Rubinstein, Ariel. 1991. "Comments on the Interpretation of Game Theory." *Econometrica* 59:909–24.
- Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge: Harvard.
- Schelling, Thomas. 1967. *Choice and Consequence*. ed. Cambridge: Harvard University Press, 1984.
- Schelling, Thomas. 1978. *Micromotives and Macrobehavior*. New York: Norton.
- Shepsle, Kenneth. 1989. "Studying Institutions." *Journal of Theoretical Politics* 2:131–48.
- Shubik, Martin. 1982. *Game Theory in the Social Sciences*. Cambridge: Massachusetts Institute of Technology Press.
- Simon, Herbert. 1979. *Models of Bounded Rationality*. 2 vols.

- Cambridge: Massachusetts Institute of Technology Press, 1982.
- Stinchcombe, Arthur. 1991. "The Conditions of Fruitfulness of Theorizing About Mechanisms in Social Science." *Philosophy of the Social Sciences* 21:367-88.
- Tugendhat, Ernst. 1985. "Habermas on Communicative Action." In *Social Action*, ed. Gottfried Seebass and Raimo Tuomela. Amsterdam: D. Reidel.
- White, Stephen. 1988. *The Recent Work of Jurgen Habermas*. Cambridge: Cambridge University Press.
- Wood, Alan. 1985. "Habermas's Defense of Rationalism." *New German Critique* 35:145-64.

James Johnson is Assistant Professor of Political Science, University of Rochester, Rochester, NY 14627.