

Alberto Vesperoni
Stockholm School of Economics
Introduction to Game Theory

Essay on “Is Language a Game?” by Joseph Heath

In this and other articles Joseph Heath discusses the Communicative Action theory of Jurgen Habermas in an original and very understandable way. In this specific paper, by discussing the work of Habermas, Heath is ultimately interested in giving support to the thesis that there exist different kinds of rationality and instrumental rationality is only one among them.

Instrumental rationality is the one we usually deal with in Economics: given preferences over outcomes and beliefs over events, a rational agent chooses actions in order to maximize his expected utility. Utility defines preferences over the outcomes, not directly over the actions. If utility would define preferences directly over actions, then we would have a different type of rationality. It would not be instrumental, in the sense of guiding the choice of the best action functionally to reaching a certain desired outcome, but would be normative. The agent would prefer a certain action over another just because he is happier performing such action, not regarding the outcome of it. A famous example of normative rationality is Kant's Categorical Imperative, that is roughly speaking the conscience of an agent telling him directly how to behave, as a goal by itself, not functionally to reaching some desired outcome. In Sociology normative rationality is the main paradigm of behavior: it identifies the discipline's methodology and it is the privileged way to explain the possibility of social order. Max Weber and Talcot Parsons are the main theorists of the social theory of normative rationality.

Habermas does not belong to this sociological tradition and has a peculiar view on the topic: he does not categorize rationality types in instrumental and normative ones, but in instrumental and communicative ones. Habermas claims that communicative behavior cannot be explained in terms of instrumental rationality, but still it is rational in its own way. Habermas claims that, for an utterance to have meaning, the sender has to be committed to justify its validity if necessary. He classifies the validity claims of an utterance in three categories: claims of truthfulness, rightness and sincerity. These three types of claim reflect respectively the three dimensions of a speech act: representational, conventional and expressive. The most important dimension for our argument is the conventional one: when an agent raises a validity claim of rightness through an utterance, he would automatically commit himself to behave coherently with respect to the implicit moral principles evoked. Habermas claims that if such a commitment would not be implicit in the utterance, then language would not have meaning. Hence, since we observe empirically that language has meaning, we can conclude that, every time an agent communicates about rightness, he is committing himself to some normative behavior. Since communicative action is based on justification criteria then it is rational in some way, but since communication always produces norm-driven actions then it has to be intrinsically different from instrumental rationality.

Based on his theory of communicative action, Habermas develops his idea of Discourse Ethics, where he claims that a verbal agreement between free persons should not need any kind of enforcement. There is no need of enforcement because the verbal agreement has been accepted by the parts and hence it is intrinsically based on a mutual commitment by them to justify their claims of rightness, behaving coherently with the moral principles they implicitly evoked.

Joseph Heath does not agree with the foundations of Habermas' Communicative action theory, and as a consequence of that he confutes the main thesis of Discourse Ethics. He developed the critique to Discourse Ethics later, in his book "Communicative Action and Rational Choice". The article I am discussing here has been written before the book, but it is the cornerstone of Heath' critique of Habermas' work. In the article Heath shows briefly the weaknesses of Habermas Communicative action theory, and then he proposes a sounder theory of not instrumental rationality, with the help of game theoretical tools.

The main point of the paper is that Heath reverses the causality in argument of Habermas about existence of language. Habermas claims that given that language is meaningful, then the claims raised by the agents involved in communicative action are necessarily valid, in the sense that the agents will be committed in justifying them. Hence, according to Habermas, language originates the ability of agents to commit in their life. Heath claims that language does not have an intrinsic meaning rooted in its semantics, but acquires meaning in the pragmatics of the real life situation in which it is used.

The idea that the meaning of language is determined by its relation with the real life situation in which it is used is not an original idea developed by Heath, but it has already a long tradition in Game theory. The philosopher David K. Lewis has been the first to formalize this intuition with game theoretical tools, modeling communication among players as a signaling game in which signals don't have any direct effect on payoffs (cheap talk). Lewis develops his idea in an incomplete information 2x2 pure coordination game where nature moves first and chooses the type of player 1, then player 1 moves, sending a signal to player 2, and then player 2 takes an action that would like to coordinate with the type of player 1. Lewis showed that there exist Nash Equilibria where the two agents are able to coordinate by using a language that both understand and expect to be understood. The strength of Lewis' idea is that meaning of language is not assumed a priori, but it develops as a convention in a completely non cooperative setting. This idea is brilliant, since it potentially explains how language could have spontaneously emerged as a natural order by repeated interaction of people, learning to coordinate the meanings by trial and error. Unfortunately the theory has two important weaknesses: the first is that the model has multiplicity of equilibria, some of which are such that language does not convey any information (babbling equilibria); the second, more serious according to Heath, is that the existence of meaning is contingent on the particular payoff structure of the game (in Lewis' model a pure coordination game). In the opposite case of a zero sum game the language would not have any meaning in all possible equilibria, because the players do not have any incentive in coordinating each others.

Vincent Crawford and Joel Sobel (1982) generalized this intuition in a continuous setting, showing that the maximum precision level of a language conventionally developed in a game is an increasing function of the complementarity of the players' interests. Personally I find this intuition very interesting, and there are many possible applications of this idea in interpreting real world phenomena. For example in "Symbolic construction of a Community", the anthropologist A. Cohen highlights how the symbol and the ideology that unifies a certain social movement will be more ambiguous the more the movement' members are in heterogeneous in goals and principles. An obvious example is how the concept of Democracy needed to be so ambiguous in the past in order to guarantee identification with democratic institutions of different social groups. The more democratic institutions boosted convergence of social groups preferences, the more democratic institutions have been defined transparently and precisely.

Because of the second weakness of Lewis' model (the contingency of meaning to the particular game), Heath claims that language cannot be modeled in a fully non-cooperative setting. Even if language could emerge as a convention in a pure coordination situation, then it would lose its entire meaning if there would exist another situation with the same actions but with incentives against coordination. Notice that, in the fully non-cooperative setting, in this case language would really lose meaning, not just credibility. From these premises Heath concludes that, in the setting is fully non-cooperative, then instrumental rationality is not able to explain human communicative behavior.

An apparently appealing solution to this problem is to assume that language has already a meaning to which people are committed to. Then, in case the players would not have incentives to coordinate their actions, language would keep its meaning and only lose credibility. The language's semantics would not be affected, only the pragmatics. The assumption that people are committed to a language meaning transforms the setting from a non-cooperative game to a cooperative one.

It seems possible to circumvent the second weakness of Lewis' model in this way, but what about the first one about existence of babbling equilibria?

Joseph Farrell (1992) developed a clever and intuitive way of excluding babbling equilibria: the neologism-proof refinement of Nash equilibria for this specific class of games. A neologism (from Greek, "new word") is a word that is not used in equilibrium, Neologism-proof equilibria is a Nash equilibria that additionally requires that there exists no neologism that would be credible and convenient to use by a player. This refinement allows excluding Babbling Equilibria, but existence of Neologism-proof equilibria is not guaranteed.

Heath claims that neologism proofness is a very intuitive property and should be required for every solution of a language game. On the other hand he claims that there are very simple games for which there is no Neologism-proof solution. From these two somehow contrasting premises he concludes that, even under the assumption of commitment to a meaningful language, instrumental rationality is not sufficient for explaining human communicative behavior.

At this point, Heath claims there is enough evidence for confuting the point of view that instrumental rationality is the only possible type of rationality that drives human behavior. He claims that language would not have a meaning if not because of norm-driven behavior of the players in language games. Hence, in his view, language gets a meaning because of preexisting norm-driven behavior of players. This view is exactly reversing the causality of the claim of Habermas, who was justifying the ability of people to commit to norms by the fact that they express their promises by a meaningful language. For Heath, communicative action does not have a special status as for Habermas, but since communication is grounded in norm-driven behavior, communication is just a subclass of norm-driven action. Hence there are only two classes of action: instrumental and norm-driven.

Heath's argument is an attempt to make sounder the foundations of Habermas' philosophical system of thought, but we can see that through this intent he is making an interpretation of Habermas' theory that is more and more sociological, bringing it closer to the approaches of Max Weber and Talcott Parsons, rooted in the Kantian philosophical tradition of categorical imperatives.

Interestingly, Stefano De Michelis and Jorgen Weibull (2008) introduced in a cheap talk evolutionary game non-instrumental rational behavior. The players in their model have lexicographic preferences for honesty: they firstly care in an instrumental way about their outcomes, but secondly they have preferences directly over their actions (they are averse to lying), that is by definition normative rationality.

Bibliography

Cohen Anthony, "The symbolic construction of a community", Peter Hamilton, 1985

Crawford Vincent, Joel Sobel, "Strategic information transmission", *Econometrica*, 1982

De Michelis Stefano, Jorgen Weibull, "Language, meaning and games: a model of communication, coordination and evolution", *American Economic Review*, 2008

Farrel Joseph, "Meaning and credibility in cheap talk games", *Games and Economic Behavior*, 1993

Farrel Joseph, Matthew Rabin, "Cheap talk", *Journal of economic perspectives*, 1996

Habermas Jurgen, "The theory of communicative action", Cambridge Polity, 1984

Heath Joseph, "Is language a game?", *Canadian Journal of Philosophy*, 1996

Heath Joseph, "Communicative action and Rational Choice", MIT Press, 2003

Lewis David, "Convention: a philosophical study", Harvard University Press, 1969