MBS 94-21
Convention As Correlated Equilibrium
Peter Vanderschraaf

A convention is a state in which agents coordinate their activity, not as the result of an explicit agreement, but because their expectations are aligned so that each individual believes that all will act so as to achieve coordination for mutual benefit. Since agents are said to follow a convention if they coordinate without explicit agreement, the notion raises fundamental questions: (1) Why do certain conventions remain stable over time?, and (2) How does a convention emerge in the first place? In a pioneering study, Lewis (1969) addresses these questions by applying noncooperative game theory. Lewis defines a convention as a Nash coordination equilibrium of a noncooperative game that is salient, that is, it is somehow conspicuous to the agents so that all expect one another to conform with the equilibrium. This paper presents a new game theoretic definition of conventions, which formalizes the notion of salience and which also generalizes the class of conventions Lewis discusses in his work. I define a convention as a correlated equilibrium (Aumann 1974, 1987) satisfying a public intentions criterion: Every agent wants his intended action to be common knowledge. I argue that many conventions correspond to correlated equilibria that are not Nash equilibria, and that this is consistent with Lewis' general viewpoint. Finally, I argue that game theoretic characterizations of convention, such as Lewis' and my own, help to explain a convention's stability, but that a fully satisfactory account of the emergence of convention requires a theory of equilibrium selection beyond the scope of Lewis' work.