Neutral Evolution as a Route to Large-scale Cooperation in the Stag Hunt Game.

Marcus Frean and Joseph Bulbulia
Victoria University of Wellington, New Zealand
marcus.frean@vuw.ac.nz

A Stag Hunt is a game with two pure equilibria: all-defect and all-cooperation. The cooperative equilibrium pays better than the defective equilibrium, and there can be no benefit from defection when all cooperate. Nevertheless, formal and empirical models agree that the defective equilibrium becomes easily entrenched. Why? The cooperative benefit ('Stag') relies on the participation of many partners, whereas the defective benefit is independent of what others do. Thus, while mutant defectors may thrive, mutant cooperators are always strongly punished. The Stag Hunt is important because it reflects evolutionary problems for human cooperation that are not captured by the more familiar Prisoners Dilemma: in a Stag Hunt what entrenches defection is the risk of losing out through cooperation, rather than the assurance of gain through defection. Various solutions to the problem of cooperation in the Stag Hunt have been proposed, but these do not scale up to large societies.

We observe that the problem of cooperation in a Stag Hunt is two-fold: some part of the system must 'escape' the stability of the defective equilibrium, and the cooperation that follows must itself be stabilized, lest the system revert. We notice that solutions which align the cooperative strategies of partners may be available wherever tendencies to express cooperative strategies may be linked to various states of the world ('correlating variables', in game theory). If the environment exists in one of several states, some of which are rare, and if we allow cooperative tendencies to randomly express themselves by coupling to specific environments, then selection against cooperative coupling will be proportional to the frequency of the states (not the frequency of Stag Hunts). In the rarest states, this pressure is near zero and the tendency to cooperate in response to them may be a random walk. We show that once coupling occurs, successful cooperation may be ratified, and become stable. Moreover, we show that if agents are permitted to adjust their environments, positive selection will favor ecological engineering such that initially arbitrary, rare states are made more common. We generalize from this model to explain how the puzzling conservation of highly affective, arbitrary symbolic environments, and the puzzling high rate of cooperation observed in large human societies are riddles that may contain each others solutions.