Comment on 'Ethics, Evolution, and Games Among Family and Neigbors' by Theodore Bergstrom.

- The object is to study four varieties of ethical (Golden) rules.
- Treating the rules as objectives one can determine the Nash-equilibria in symmetric two-player games.
- The results relate the Nash-equilibria under various rules.

- The Golden rules are associated to Hamilton's rule in evolutionary theory via associating the rules with fitness.
- The ESS and the Nash-equilibria of symmetric two-player games with rules as objectives are related.
- Results for asymmetric games are given via an evolutionary interpretation where players may be in two different roles.
- The relations between the Golden rules and evolutionary fitness criteria are established.

- My comments pertain to the first part of the article; whatever 'problems' one finds there have repercussions to the fitness criteria.
- Another approach to ethical rules would be to determine them as equilibrium outcomes of repeated situations.
- Now the individual pay-offs of the games do not correspond to the individuals' preferences.
- I presume that this route retains the independence of the choices; otherwise one could also play with the choice sets or the games.

- Consider the ethical rule 'maximise your opponent's individual pay-off'. In the prisoners' dilemma the Nash-equilibrium would be (co-operate,co-operate).
- One could model this also as a game where Player-1 chooses from Player-2's choice set and vice versa.
- Or one could consider a game where the pay-offs are given by $v_i(x_i, x_j) = u_i(x_j, x_i)$ where u_i is the utility function of the original prisoners' dilemma.

- My main question pertains to how complete the modelling should be.
- The Golden rules are advocated by religions which compete for followers.
- Presumably the choice of relegion/Golden rule should be modelled as an equilibrium phenomenon.

- Or should one think that there is some level-2 Golden rule that is used to select between Golden rules (level-1 Golden rules).
- If one goes this way it raises an interesting question whether any of the Golden rules has the property that it selects itself from amongst a set of competing Golden rules.
- The incomplete modelling manifests in the Negative do-unto-others rule in the redistribution game.
- Clearly, neither agent regards outcome $(\$100, \$20)$ nor outcome $(\$20, \$100)$ as desirable.
- So what keeps the agents from continuing their relationship and implementing a further redistribution?

- The second issue I raise is that many interactions between people involve more than two persons.
- How do the Golden rules generalise to n-player situations.
- Consider a game of private provision of a public good.
- The cost of contributing x_i towards the public good is $-x_i$.
- Given that the other agents j contribute x_j the utility of agent i is

$$log\left(x_i + \sum_{j \neq i} x_j + 1\right) - x_i$$

- Let us determine the symmetric Nash-equilibrium.
- Assume that agent 1 contributes x and all other agents contribute y. Agent 1's problem is to maximise w.r.t. x

$$log\left(x+(n-1)y+1\right)-x$$

- From the first order condition evaluated at $x=y$ one gets $y=0$ as the equilibrium contribution.
- The socially optimal level is given by maximising w.r.t. y

$$nlog\left(ny+1\right)-ny$$

and the solution is $y=\frac{n-1}{n}$.

Love thy neighbour as thyself.

Bounded love budget.

- One maximises w.r.t. x

$$log\,(x+(n-1)y+1)-x+\frac{1}{n-1}\sum_{i=2}^{n}\{log\,(x+(n-1)y+1)-y\}$$

and finds that in equilibrium $y=\frac{1}{n}$.

Unbounded love budget.

- One maximises w.r.t. x

$$log\,(x+(n-1)y+1)-x+\sum_{i=2}^{n}\{log\,(x+(n-1)y+1)-y\}$$

and finds that in equilibrium $y=\frac{n-1}{n}$.

Do-unto-others rule.

Bounded love budget.

- If everyone else contributes y and player 1 x, he acts as if he chose y and all others x, and considers their pay-off maximising w.r.t x

$$\frac{1}{n-1}\sum_{i=2}^{n}\{log\,(y+(n-1)x+1)-x\}$$

  and finds that in equilibrium $y=\frac{n-2}{n}$.
- Unbounded love budget gives the same result.

Kantian rule yields the socially optimal solution.

Negative do-unto-others rule may be difficult to formalise in general n-player settings as x_0 is most probably a function of others' choices; it may not exist.

Finally a small comment.

- Have we any reason to expect that when people's preferences are given by one of the Golden rules, total welfare maximisation has any relevance? Or do we know how to aggregate utilities?