0.0.1 Rational Choice Model

The rational choice model is similar to the socialized actor theory in that they both assume that there is a rational basis for people conforming to norms. The two theories differ about what the rational basis is. While the socialized actor theory focuses on guilt and self conflict associated with acting in ways that do not conform to social norms as the motivation for conforming, the rational choice model states that social sanctions imposed on non-conformers by other members of society constitute the primary cause for conformity [**?**]. The sanctions might take on different forms such as shaming, ostracism, or in the case of Donald Trump's rallies, punches administered to protestors. Since no one wants to be subjected to such sanctions, it is more beneficial to conform.

Clearly, sanctions can only be enacted against observable behaviors. Thus, this is a very behavioral approach to thinking about social norms and proponents of this approach have mostly been interested in why people coordinate their efforts, particularly in the equitable and sustainable use of public goods. Due to its emphasis on rational and self-centered actors, it is also easy to adapt this model to cost-benefit analyses. For these reasons, economists, political scientists, and philosophers have been very attracted to this way of thinking. Brian Skyrms, Peyton Young, Cristina Bicchieri, and many others have used game theoretic methods based on this model to explore questions of coordination in the use of common pool resources and to explore the evolution of social norms more generally [**?, ?, ?**].

In modeling the dynamics of norms using this line of thinking, the tendency has been to define individual payoffs from either cooperating or defecting in terms of money. Experimental evidence has not supported this definition of utility. A number of studies have shown that even if utility maximization is the driving force for conforming to social norms, monetary values do not serve as good proxies for utility [**?**]. Ernst Fehr has argued vehemently against using game theoretic approaches based on monetary utility functions by employing a number of different lines of evidence that cast doubt on the assumptions that these models make.

Under the rational choice model, one of the principal requisites for social sanctioning to be effective is repeated interactions between the individuals comprising a society. If there is very little probability of two people ever meeting again, there is little incentive for them to cooperate when cooperation is costly. There is also very little incentive to sanction cheaters in one-off interactions because the sanctioners gain nothing even if the imposed sanctions make the cheaters change their behavior in the future. However, Fehr and his collaborators have shown through a series of clever experiments that people do cooperate and that they do punish cheaters even in interactions where there is a negligible possibility of interacting with the same person again. In fact, people punish cheaters even if the actions of the cheaters do not directly affect them, but instead affect some unrelated and unknown third person [**?, ?, ?**].

Fehr stresses and shows the importance of social sanctions to establish and maintain conformity in his experiments; but he also shows that sanctioning is

observed even if it is costly and even if the sanctioner has nothing to gain from the act of sanctioning. Her actions are not visible to other individuals, so an increase in social reputation cannot be used as an explanation either. It might be that norms of fairness and justice are internalized in most people to a degree such that not punishing violators produces feelings of guilt and self reproach. Experiencing such negative emotions could be costlier than bearing the expense of sanctioning the violators, even if the expense to the sanctioner is considerable. In such a case, even though the motivation for sanctioning is selfish and rational, quantitatively comparing costs and benefits becomes difficult.

One way to get around this problem has been to assume that norms of justice and fairness already exist in the community being modeled [?]. This transforms the outcomes so that payoffs resulting from cooperating are greater than from not cooperating. As is obvious however, such a model cannot possibly explain the emergence of all norms. How could the norms of fairness and justice have originated in the first place under such a system?

Also, many of these models play out as pairwise interactions between individuals. However, most real world situations involve many people, all of whom have to coordinate their efforts in order to achieve some common goal. When the game theoretic approach is extended to multi-person interactions, the conditions under which conformity is established and maintained turn out to be very restrictive [?]. For example, when the number of individuals is large, achieving conformity is nearly impossible. When the number of individuals is small, the model dynamics closely approximate the pairwise interaction models. It seems to me then such game theoretic approaches are not going to be very useful when applied to the problems that I am most interested in.

Additionally, there is evidence to suggest that social sanctions are not the only way of ensuring conformity. A.S. Diamond, in his book *Primitive Law*, cites examples of many cases in the tribes he studied where norms were not upheld by sanctions [?]. Also, as mentioned in the previous section, group membership can also be very effective for establishing and maintaining conformity. In fact, group membership, even in the absence of any possibility of social sanctions, can produce a high degree of conformity and cooperation [?]. Rationalists often try to explain these observations in terms of utility as well. For example, one of the explanations proposed for why group membership might ensure conformity is that once someone has wilfully identified with a particular group, violation of any of that group's norms would be very costly because it would result in a lowering of his self-esteem [?]. While one might be able to couch these alternate mechanisms in utilitarian terms, it is much harder to formulate these arguments into explicit utility functions. It is also improbably that these utility maximization based arguments tell the whole story. Modeling people as purely rational agents is extremely unrealistic. While these rationality based models might conform to reality some of the time, it is hard to imagine that a model arising out of such restrictive assumptions will be of much use in predicting human behavior more generally.

Finally, one of the big issues with the rational choice model comes from the observation that sanctioning could at times be costly. Why then would a

perfectly rational individual ever want to sanction someone else, especially as a third party? In order to address this problem, rationalists invoke the concept of metanorms [**?**]. Metanorms are second order social norms that instruct members of society to sanction individuals who do not punish violators of other norms. This line of thought can be carried on *ad infinitum* however, and in order to achieve conformity, society would need to have infinitely many levels of norms, each level punishing violators of the level immediately below. Additionally, Fehr's work has shown that people will punish defectors even when they themselves are not subject to any social sanctions whatsoever [**?**].

To me, the rational choice model is probably the least useful way of thinking about conservation related norms. It has been shown that in order for social norms to be effective, rewards for conformity or punishments for violations have to be immediate [**?**]. Unless there is some external incentive provided, there is no tangible immediate benefit to people conforming to conservation friendly norms. Benefits, if any, will be far into the future and will not accrue to the conformers.

The environmental movement faces similar problems, and efforts have been made to make the benefits of pro-environmental behavior more personal and immediate. For example, in one of the more famous norm interventions of our time, a utility company called Opower started including the average electricity consumption of relevant neighborhoods in customers' monthly electric bills [**?**]. If a particular customer's energy usage was less than the average, she would get a smiley face on her bill. This simple intervention reportedly led to close to a 2% reduction in monthly energy consumption. Of course, this raises a number of important questions about what happens if the incentive is taken away. Does consumption return to pre incentive levels? Does consumption increase because earlier motivations for using less energy get replaced by the provided tangible incentive? How long does it take for incentive induced behavior to become a norm, either personal or social? What incentives are best at affecting change? There are no clear answers to any of these questions.

Given that people often act in ways that are beneficial to their identified group members, for reasons that are not yet well understood and that are not fully explained by pure rationality, maybe a better approach to making pro-conservation behavior a norm would be to frame policies in a way that allows people to think of members of other species and of nature in general as part of their communities. The 'nature for people' approach to conservation has tried to do precisely that, but the role that this approach assigns to nature and biodiversity is a socially precarious one. The role is that of a resource or a servant to the larger human enterprise, not as a member enjoying any degree of equality with the human part of the community. While I do think that the ecosystem services approach is a step in the right direction towards a more inclusive view of the relationship between humans and nature, it gives rise to a whole host of other problems (as discussed in section **??**). I am much more in favor of the 'humans and nature' philosophy, but on a much larger scale than the one on which it is presently being implemented on.

I hope that by this point, I have convinced you that we need a more ef-

fective way of doing conservation and that social norms might provide us with this effective new way. I also hope that I have convinced you that we do not understand norms well and that we need to understand them better in order to be able to make use of their considerable power. In the next chapter, I will inform you about my ideas for exploring whether changing social norms can lead to any positive conservation outcomes.