Social Approval as a Selective Incentive and a Collective Good: The Role of Normative Motivation

Abstract

The following essay explores two models of using social approval in sociological rational choice explanations. The first, in which social approval is modelled as a collective good, is associated with the work of Michael Hechter. The second, in which social approval is modelled as a selective incentive, is associated with the work of James Coleman. In Hechter's model, the emphasis is on formal, centralized approaches to producing collective goods. In Coleman's model, the emphasis is on informal, decentralized approaches to cooperation. The argument presented here focuses on the failure of what is called the "standard model" of rational choice, in which self-interested wealth maximizing egoism is assumed. Further, the standard model treats normative beliefs as non-intrinsic. The argument here shows that neither Hechter's nor Coleman's explanatory aims can be achieved by using the standard model that they employ. I show that my criticisms are related to motivational crowding theory in economics. I conclude that any model which uses social approval as a selective incentive or as a collective good must also employ some notion of intrinsic normative motivation.

Keywords: Normative motivation, Formal controls, Social approval, Motivational Crowding, Cooperation

1. Approval Incentives in Sociological Models of Collective Action

The aim of this essay is to provide an examination of how social approval is modelled as an incentive in sociological rational choice models. I will examine how social approval is modelled as an incentive that is itself (at least partly) the aim of a group, that is, where the collective production of social approval is the collective good in question. I will secondly examine the use of social approval and disapproval[1] as a selective incentive or disincentive in eliciting cooperation (where "cooperation" means cooperation in mixed-motive 2 or n-person games, which is known variously as the problem of public (or club) good provision (in economics), collective action problems (in sociology) or social dilemmas (in social psychology). I will illustrate my claims largely negatively by criticizing previous models of social approval as they are discussed by sociologists in the rational choice tradition.

By way of introduction, I will here simply list the claims that I will make in this essay, and then briefly discuss how I will proceed.

1.1 The assumptions of the argument

The claims that will be employed in this essay are as follows:

[1] Henceforth, any claims made for approval hold for disapproval, and vice versa.

C1: Actors who cooperate and who punish failure to cooperate are motivated by a belief, attitude or value that some action is right or good[2], and not *solely* by material interests.

C2: Actors also cooperate in pursuit of material interests, but subject to constraints on normative beliefs. [3]

C2a: Normative beliefs about justice may include: (a) equity, where justice suggests that the rewards of cooperation should be proportional to inputs; (b) equality, where justice suggests that rewards of cooperation should be divided evenly amongst a group's members; (c) need, where justice suggests that people in special circumstances should receive a portion of the rewards of cooperation even if they did not cooperate (which conflicts with equity), or that they should receive more than a fair share of the rewards of cooperation (which conflicts with equality).

C2b: Other normative beliefs may include beliefs about certain virtues or action descriptions in which actions may be seen as loyal, friendly, kind, trustworthy, brave, and the like.

C2c: Normative beliefs include a concern for the fairness of procedures and not just outcomes.[4]

---

[2] For purposes of the argument as a whole, I will use the term *normative belief* instead of *attitude* or *value* or *moral desire.* I will refer to what an agent values to be either right or good, although I will make distinctions between what Rawls (1971) calls "the right" and "the good" when I discuss procedural fairness in section 1.1 (which is part of what Rawls means by "the right"). All my position requires is that the actor is capable of being persuaded by argument that appeals to normative factors, not unnamed forces which induce conformity or imitation.

[3] That fairness *in distribution* can be distinguished into the three categories of equity, equality and need are discussed by Deutsch (1985) in a social psychological context and Miller (1976) in the context of political philosophy.

C3: Moral motivation is heterogeneous, both between and within individuals. That is:

C3a: Different actors are motivated by their normative beliefs to varying degrees, and

C3b: Actors may themselves be motivated to act on their normative beliefs to different degrees (see C5 below).

C4: Those actors who do not cooperate are capable of being persuaded to do so.

C5: In order for non-normatively motivated actors to be treated as rational in being persuaded, and in attributing rationality to those who persuade non-normatively motivated actors, we must make clear what is going on "inside their heads" (i.e., we must attempt a "rational reconstruction" of their actions).

C5a: Normatively motivated actors believe either that (1) they have an obligation to sanction non-cooperators[5] regardless of the efficacy of the sanction, or (2) that those who receive sanctions are capable of changing their minds about what action is normatively required.

C5b: Non-normatively motivated actors, if they respond to a sanction do so because they (1) feel embarrassed about being sanctioned, (2) they feel shame because they have done something that they believe to be wrong.

C6: Shame, in contrast to embarrassment, operates through one of two mechanisms, which correspond to two different theories of weakness of will (or of wrong-doing in general). One is the knowledge account: an actor does something wrong because he does not know what is right. A sanction induces shame because the actor feels that he should have *known* what was considered right. The second is the motivational failure

---

[4] See Lind and Tyler (1988)

[5] Assume that matters of factual belief are held constant. Namely, the probability of (a) the contribution of others, and (b) the level of need of others, is common knowledge.

account: an actor *knows* what action is the morally right one to take, but fails to be *motivated* by that knowledge.[6]

1.2 Illustrating the claims

While the answer to which version is the correct one from C6 is philosophically contended, all the other claims should be readily accepted, and can be accepted regardless of which interpretation of wrongdoing is favoured. I will be concerned primarily with the motivational failure account.

In this essay I will examine two attempts to use the notion of social approval to explain cooperation and solidarity using what I call the "standard model" of rational choice, by which I mean egoistic maximizing behaviour. The standard model not only makes motivational assumptions, it also rules out alternative models of utility maximizing behaviour which describe systematic departures from the standard model, such as framing effects, bounded rationality, myopic decision making, and the like. It also excludes emerging areas of research into types of normative motivation in the social sciences at large.[7]

I choose two such representative developments of the standard model, those presented in books by Hechter (1987) and Coleman (1990). There are two reasons for focussing

[6] The position that one does wrong and later regrets the action because he did not know the right action to take when acting (and not because of a motivational failure) is known in the philosophy of action as specificationism. That is, an actor has yet to specify what action he thinks to be right, and hence can act "wrongly", but not in a way that the agent perceives to be wrong (see Millgram (1997: 135-8)).

[7] In economics see Fehr and Schmidt (1999) and Bolton and Ockenfels (2000).

on these two books as representative of sociological rational choice approaches. First, and most importantly, Coleman and Hechter represent two different approaches to modelling social approval using rational choice models. Hechter attempts to *explain* how social approval is produced as a collective good. Coleman, in contrast, *assumes* social approval as a selective incentive to produce other public or collective goods. I will argue that both fail because of their use of the standard model. In particular I will argue that each, in different ways, is subject to a criticism from recent economic theory on motivational crowding.

The second reason for focussing on these texts, discussed extensively below, is that they represent two different poles in explaining collective action and the emergence of norms. Hechter's approach is based in the establishment of formal controls and centralized organizational structures to maintain norms. Coleman's approach, conversely, is in the tradition of explaining norms from an informal, decentralized approach. That each of these very different explanatory traditions are subject to the same criticism should be instructive to followers of either tradition.

The third reason for focussing on these texts is that no other works in sociology have been as systematic in their approach to traditional sociological problems using a rational choice approach. To examine their respective faults should show the limits of applying the standard model of rational choice to sociological problems of group formation and maintenance. Further, both of these books have been extremely influential as representative positions of the rational choice approach within sociology, and if there are faults within their positions, as I shall argue, it is all the more important that they be recognized.

2. Social Approval as a Collective Good: Hechter on Intentional Goods

In this section I first introduce Hechter's general theory of cooperation in the production of collective goods as being the result of formal organization. I then criticize the assumptions of his theory as applied to the production of what he calls intentional (immanent) goods, which are analogous to collectively produced social approval.

2.1 Hechter's Model of Solidarity: External Mechanisms for Cooperation

Hechter's rational choice contribution is more in the economic style of transaction cost economics (Williamson, 1985) or new institutionalism (North, 1990) for a general introduction), in that he focuses on the need for centralized, formal organization to explain production, in contrast to the pure market approach of organizing production through a system of prices. That is, Hechter's solution to the problem of the emergence and maintenance of norms relies on the production of sanctioning systems by external agencies such as organized groups, firms, and states. This provides a helpful contrast to the more "internal" or decentralized approach taken by Coleman and discussed in section III of this essay.

Hechter's *Principles of Group Solidarity* (1987: hereafter, PGS) seeks to explain how groups form, and then develop and maintain solidarity. Solidarity for Hechter is a continuous variable defined as the degree of compliance with norms prescribed by the group and the scope of those norms (where scope refers to the extensiveness of the

demands of the norms). Groups are formed because of the need to produce joint goods, which are goods that require other actors to be produced and are excludable to non-group members. Thus, collective goods are more easily produced than pure public goods because they can exclude non-contributors (i.e., non group members) and can avoid free riding from outside the group, but must face a free-rider problem within the group.[8]

Further, where selective incentives are used to explain the formation of groups and the provision of collective goods, Hechter (rightly) argues that the provision of selective incentives is itself a collective good. That is, for one actor to sanction a non-contributor (a selective dis-incentive for the non-contributor), the sanctioning actor himself requires that there be a selective incentive for him to produce the selective (dis-)incentive. Call this the second-order collective action problem. Then, of course, there must be a selective incentive for actors to provide a selective incentive for those actors who supply the first selective (dis-)incentive, and so on. The second order collective action problem is sometimes referred to as the meta-norm problem, and I shall call the subsequent regress in the production of selective incentives the *metanorm regress problem*.

In Hechter's theory the metanorm regress problem is resolved because individuals commit themselves to a group whose aim is to produce a collective good, and agree to

[8] Hechter is here following Buchanan's (1965) pioneering discussion of the economic theory of clubs.

establish a system of monitoring performance and sanctioning or rewarding[9] non-performance in order to produce the collective good. On Hecther's view then, individuals can only produce a collective good by committing to form a group which has the power to force that set of individuals to produce the good.

Hechter's theory is what he calls "contractarian" (1987) or "solidaristic" (1991). The relevant features of his theory are the following: (1) it relies on intentional design, versus an invisible hand or market price system of producing collective goods. (2) A group thus intentionally creates obligations which members must fulfil in order to receive their share of the collective good. (3) In order to ensure compliance with obligations, groups can either (a) compensate members for their effort, or (b) reward the members through the provision of what Hechter calls "immanent goods". Immanent goods are those goods which are non-compensatory, and "directly satisfy their members' utility" (PGS: 42). It is the reliance on immanent goods that distinguish solidary groups from groups (such as firms) that rely on material rewards to ensure compliance. (4) The level of a group's solidarity, then, is a function of (a) the extensiveness of its obligations, and (b) the degree of compliance with those obligations. (5) The degree of compliance in turn is a function of (a) the overall level of dependence of members on the immanent goods, and (b) the degree of monitoring and sanctioning that the group has at its disposal (Hechter calls (b) the "control capacity" of the group).

[9] Hechter uses sanction to refer to a positive or a negative sanction: I will stick to ordinary language and refer to a sanction as a punishment or a source of disutility for non-performance, and call a positive sanction or source of utility for performance a *reward*.

Now, because Hechter's goal is to show that rational choice theories can explain group solidarity in a more satisfactory way than its sociological rivals of structuralism and normativism, Hechter stays close to the standard model. However, as opposed to the pure wealth maximizing of the standard model, Hechter's contribution to sociological rational choice is to assume egoism in the realm of the pursuit of non-material goods. That is, actors will want goods that are provided collectively (i.e., not from the market), and which can include social goods, but not want to produce them. As stated by Hechter: "whenever people are faced with two divergent courses of action- one in pursuit of some individual end, the other in pursuit of some collective end- I will assume that they will invariably choose the former" (PGS: 37). So, in the case where an actor may desire some "social" good (i.e., approval), he is still unmotivated to act to get the good if he can get that good without effort.

Note, then, that Hechter's account does not rely at all on the idea of social pressure or normative expectations of compliance. Even though there may be "informal" sanctioning, this would only be motivated by the existence of a formal agency who could reward the individual for providing social pressure, and this is the actor's only motive according to PGS. The view of motivation in PGS is that there is always a better explanation of cooperative behavior which is based on egoism. This is illustrated in Hechter's challenge to the normativist explanation of social sanctions in rotating credit associations:

> In many societies the state will prosecute defaulters and make them subject to imprisonment. But the literature on rotating credit associations suggests that other types of sanctions are more likely to come into play. Geertz (1966), for instance, claims that the incidence of default in rotating credit associations is rare because the members are frequently fairly close acquaintances and so would be deeply ashamed to evade their obligations. Members are said to go to great lengths, such as stealing or selling a daughter into prostitution, in order to fulfil their obligations to the association…Fortunately, a much simpler way of explaining community sanctioning is at

> hand. A person who is known to have defaulted in one rotating credit association is not likely to be accepted as a member of any similar group. (PGS: 109).

Now, this justification for the "simpler way of explaining community sanctioning" has no force whatsoever. First, the account based on future membership in other rotating credit associations is not *simpler* at all. Both the normativist account and the egoistic account only relies on an actor having one aim: the desire for social approval and the desire for future income, respectively. Just because many see the desire for material gain as more basic than the desire for social approval does not make it so. Second, while it is certainly possible that the normative explanation can conceal an egoistic motivation, why this is taken as the only possible state of affairs is undefended.

What Hechter and the proponents of the standard account ask social scientists to do is to hedge their bets in favour of egoism: while some actors may be normatively motivated, most normative motivation in fact is a rationalization of some form of egoistic maximizing. But this argument is speculation and assertion, without argument. I will show that in fact when the argument is further examined, it in fact fails to be logically coherent.

2.2 Criticisms of Hechter

2.2.1 Can Egoists Make Agreements to Act Collectively?

Despite the fact that I believe that one needn't adopt the egoistic motivational assumptions to explain group solidarity, I believe that the argument fails on its own terms. There are two basic criticisms with which I shall begin.

The first is that Hechter's theory exhibits inconsistencies in terms of the resolution of collective action problems through group formation. That is, by arguing that groups overcome collective action problems by forming groups that impose obligations, Hechter sidesteps the problems that any group of egoists will have in coming to make agreements. For, the crucial step in Hechter's argument that groups can provide collective goods relies on the ability of individuals to secure credible commitments to enforce agreements to enforce agreements that the individuals know that they do not want to fulfil. That is, an actor wants on the one hand to both have the good and hence must resolve to do his part in collective action along with others; on the other hand, he wants to receive the good with minimal (and preferably no) effort.

The problem of group *formation*, then, is how an egoist can ever come to signal his willingness to produce a collective good such that an agreement can be made. This problem is initially met with the problem of revealing demand for the collective good (what Hechter calls the individual's level of dependence). Hechter, wrongly as I shall argue, sees the problem as follows: "The voluntary establishment of formal controls in non-hierarchical groups is not difficult to explain in rational choice logic. What *is* difficult to understand is why the rational members of large groups would ever abide by the controls they have consented to establish." (PGS: 123-4; emphasis in original).

The point here is that Hechter fails to acknowledge that the second problem of abiding by controls relates to the first problem of how the group members consent to establish controls. For, if individuals clearly do not want to act so as to produce a good, and this egoistic rationality is common knowledge, then no one can actually rationally

consent to the controls, as he knows they will not be enforced. Thus, if each individual knows that all are egoists, then only a solution assuming some non-standard motivations will work to explain the formation of the group, or any subsequent compliance with the group's obligations. This may be seen as a variant of the general problem of credible commitment in securing cooperation (for the theory of credible commitments, see Schelling (1960)).

2.2.2 Can All Immanent Goods be Produced by Egoists?

Another inconsistency in the application of standard rational choice assumptions is more damaging to Hechter's argument than the credible commitment problem, a problem which goes more to the heart of the problem of assuming egoism. My argument is that Hechter must assume certain normative motivations to produce a type of immanent goods, namely social approval goods.

To see why Hechter's motivational model (i.e., of dependence) implicitly relies on some normative motivation, I will examine more closely the notion of immanent goods. Immanent goods are those for which, simply put, one expects some reward in the future, and/or the utility is not provided by compensation. However, the nature of immanent goods is not entirely straightforward. I will therefore class the examples of immanent goods discussed by Hechter into three relatively discreet categories. There are first goods which can be classified as what economists refer to as "delayed payment contracts". These include: "insurance, welfare benefits, and the prospect of a

career" (PGS: 142), where the existence of a strong internal labour market in a firm or mutual aid societies would be prime examples. The second type of immanent goods in the proposed scheme are goods which are essentially consumable goods that could be provided by the market: "sense pleasures, happiness, and so forth." (PGS: 42).

While not the sole focus of my critique, it is worth making a brief comment on this second type of immanent good. The criticism is as follows: if goods which are readily available on the market are instead sought through solidary groups, then it must be that the solidary group provides some *other* good as well, and presumably some good which is not available on the market. An example of the type of good discussed here could be membership in the cooperative societies in the U.K., and other types of cooperative societies whose aim is to provide typical market goods, but do so according to principles of organization which differ from those generally found in firms who exist on the competitive market. These groups essentially provide market goods, but organize their method of distribution and control in a different manner from most firms. If this analogy is correct, membership may be quite diversely construed, such that merely buying from an "ethically conscious" store is no different from being a member of a cooperative which provides market goods.

The third class of immanent goods are those that Hechter calls immanent goods which are produced in intentional communities, or what Pettit (1990: 740) calls "attitude dependent" goods. It is this case that is the focus of my critique. These are goods, defined by Hechter, such as "the social contact engendered in cooperative labor" (PGS: 141), "…a sense of community, friendship, love, and the feeling of security – all of which flow from the existence of social harmony" and "love and friendship"

(PGS: 171). Now, recall that in Hechter's framework, the assumption of egoism or self-interest, is not that individuals are solely oriented to attaining and consuming material goods, but that individuals will choose what will benefit themselves before any goal of a collective, and individuals will always prefer that someone else besides themselves produce the good in question. However, this logic of egoism becomes incoherent when applied to social goods. This is because the nature of enjoying the consumption of friendship requires that one produce it as well. That the good be produced and consumed communally is a feature of these attitude dependent goods.

To see the case made here, a simple example will do. Imagine that actor A says to B: "I would like to have your friendship; however, I would rather not give you any of my friendship." Such an approach to social approval would (a) be unlikely to get many offers of friendship, and (b) seems to belie the very logic of friendship. That is, to want to have a friend or enjoy the good of "community, friendship, love, and the feeling of security - all of which flow from the existence of social harmony," would seem to require that one is willing to contribute to the social harmony itself. Now, it may be that there are some forms of perverse friendship or types of social harmony that I am ignoring, but given that these forms deviate so substantially from the norm, my point still holds.

Note further how this logic of friendship applies to the group production of social harmony. It would require that when a member of an "intentional community" fails to produce his share of social harmony, he would be compelled to do so. This approval to be consumed by the collective would hardly seem to be worth much, if it has to be

coerced out of the actor[10]. Like the case of the friendship where the desire is to get friendship without producing any, the logic of free riding in the production of social goods fails to make sense of the normal practice of friendship.

The preceding argument suggests that to organize the collective production of social approval, what I have called intentional immanent goods or attitude dependent goods, is an inefficient means of producing it. Thus, the argument that social goods can be produced through the most extensive system of monitoring and sanctioning, Hechter's primary argument, does not hold for the production of intentional immanent goods. Indeed, one can argue the opposite: that the greater the degree of coercion, monitoring and sanctioning, the *less* likely it is that social approval will be produced. On this view, what is to be explained is how groups with an aggressive approach to promoting sociality manage to remain in existence. Casual observation would suggest that most cults or communes tend to not last as long as, say, mutual aid societies (assuming that their control capacity is equivalent). Nevertheless, Hechter's account is
My conclusion is that in regard to the production of social approval as a collective good, the rational choice account on offer from Hechter is logically fallacious and empirically unsupported.

This claim can be made in the language of motivational crowding theory, largely developed in economics by Frey (1997; Frey and Jegen, 2001): to produce social approval one must be *intrinsically* motivated to produce it. Intrinsic motivation is contrasted with extrinsic motivation, where an actor requires some external

[10] As put by Offer, writing on the "economy of regard", "…to have value, regard must be authentic, i.e., unforced." (Offer, 1997: 454)

intervention in order to produce a good. External intervention may be monitoring performance, providing incentives for performance or sanctions for non-performance. Now while it is always the case that there must be some incentive to undertake an effort to produce, some incentives are intrinsic to the activity itself. For example, in the production of "social harmony", it may be the case that one is offered financial reward for liking one's fellow group members. But such social approval would be of little value. Instead, the incentive required to produce such a good is one that is intrinsic to the activity itself; that is, to behave in a kind, friendly manner to other people just because he believes they deserve to be so treated, not because of the prospective kindness and friendship one wants to receive in return. Although it is the case that one may enjoy receiving approval in return, it is not the reason one behaves kindly; if it is the reason, one is unlikely to receive approval in return. To receive friendship from others is, as Elster calls it (1983: ch. 2), a state that is essentially a by-product. That is, one must not aim at receiving a good in their actions in order to receive that good.

I will show that it is this same error of ignoring the intrinsic nature of social approval that leads to the failings in Coleman's account of the emergence of norms, to which I now turn.

3. Social Approval as a Selective Incentive: Coleman on the Emergence of Norms

As stated in the introduction to this essay, Coleman's account of the emergence of norms contrasts with Hechter's in that Coleman believes that norms can emerge without the use of formal organization. Instead, Coleman argues that social approval

can act as an incentive in the emergence of norms. That is, Coleman argues that social approval can lead to a "moral code" and the internalisation of norms. I will show that in fact a moral code or sense of justice is required to make sense of social approval, and hence social approval cannot precede normative motivation in the order of explanation.[11]

Coleman's *Foundations of Social Theory* (1990: hereafter, FST) seeks to explain social facts such as the "genesis and maintenance of norms, adherence of persons to norms, development of a moral code, identification of one's own interest with the fortunes of others, and identification with collectivities" (FST: 31), which must be explained from the assumption that actors are "norm free, self interested" (ibid). In this section I will show that these social facts (normative systems, moral codes, compliance with norms) must be assumed and cannot be explained from premises of self interest.

3.1 Approval and Normative Sanctions[12]

[11] A related argument can be found in Mansbridge (1997).

[12] Coleman uses the term "normative sanctions" (cf. FST: 292) which is actually more appropriate than "social disapproval" for my purposes. That is, it emphasizes that there is a normative content to a social sanction which has effect for an agent, not merely that an agent expresses his dislike for an agent because that agent's actions are not normatively sanctioned. The problem arises, I will argue, in that the agent sanctioned must have some understanding of the normative content for the sanction to be effective, and this normative understanding is what the account of normative sanctions *is itself meant to explain.*

I begin by describing the general structure of Coleman's theory. For Coleman, the aim of social theory is to explain how norms are created and maintained. Thus, Coleman sees himself as addressing the problem of social cost, or how actors can reduce the degree of negative externalities in interaction and exchange through establishing norms. Norms have focal actions which can be prescribed or proscribed, and the target of the norm is that class of individuals of whom the behaviour is expected, and the beneficiaries of the norm are those who have an interest in the reduction of negative, or the creation of positive, externalities. Norms are conjoint when the targets and beneficiaries of the norm are the same class of actors, and disjoint target and beneficiary are separate actors.

For a norm to emerge, actors must have a *demand* for the norm (i.e., a need to reduce externalities), and a means of enforcing the norm. This brings up the problem of the second-order collective action problem or metanorm problem as discussed above. Norms will emerge when "…beneficiaries of a norm, acting rationally, either will be able to share appropriately the costs of sanctioning the target actors or will be able to generate second-order sanctions among the set of beneficiaries that are sufficient to induce effective sanctions of the target actors by one or more of the beneficiaries". (FST: 273). But what overcomes the metanorm regress problem? Here Coleman suggests, in contrast to Hechter's emphasis on formal controls, that the motivation to sanction others comes from the support provided by close social relationships (i.e., what Coleman calls networks under conditions of closure (FST: 275-6), or low exit opportunities, where "zeal" (explained below) is possible).

However, this motivation only comes about because the closure of the social network, or the formation of the group, is already assumed. Of course, a sociological rational choice approach founds itself partly on the idea that the formation of groups is itself a collective action problem (if only a coordination problem) that stands in need of explanation, and this is Hechter's contribution. Coleman states that the metanorm regress problem is solved by assuming a desire for approval and a fear of disapproval. Coleman calls this the "rationality of zeal": "The rationality of zeal has the same incentive that leads to free riding, but with a second incentive superimposed on the first. The second incentive, however, becomes effective only through an intervening action: encouragement of others, or positive sanctions, which may overcome the deficiency of the first incentive". (FST: 275).

Let me explain what Coleman means in the previous passage. The rationality of zeal is the second-order collective action problem: that of there being a rational incentive for punishing actors who do not cooperate. As stated, the second order collective action problem is that there needs to be an incentive to reward those who punish non-cooperators. This is the second incentive to which Coleman refers: the motivation to sanction non-cooperators through the use of social disapproval is motivated by the social approval that the sanctioner will receive from others.

There is a seemingly obvious problem here. The nature of the collective action problem is that there is an incentive to free ride. To solve this collective action problem free-riders must be sanctioned, but the sanctioner must too be rewarded. And how is the sanctioner rewarded? Through social approval, in Coleman's account. But note that Coleman has assumed at the second-order level precisely what he seeks to

explain at the first-order level, that is, the motivation of social approval to promote cooperation. But why does this itself not lead to another level of collective action problems (i.e., a third-order collective action problem), leading to the metanorm problem I have described? Coleman here simply assumes away the second order problem by assuming that the third-order problem is solved.

In FST (271-2, 821, 926-30), Coleman suggests that if interests are sufficiently high amongst certain actors in the production of a norm, then they have an incentive to sanction. That is, if the benefits to the introduction of a norm are perceived to be greater than the costs of sanctioning to an actor, then he has an incentive to sanction non-cooperators. As Coleman argues "(t)his cost reduction to norm beneficiaries may give them an interest in establishing a sanctioning norm" (FST: 273). However, the degree of interest or the cost of sanctioning has no bearing on solving the higher-order collective action (i.e., metanorm regress) problem, nor does it explain the efficacy of sanctioning given the paucity of its egoistic content for inducing cooperative behaviour, as I shall argue.

Now, recall that the motivational assumptions employed by Coleman are that actors are not motivated by norms, and are self-interested. It is this that gives rise to the free-rider problem in the first instance. For a moment, set aside the metanorm regress problem that I have discussed, and instead focus on the intentional content assumed by the assumption of self-interest in motivating actors to sanction and in motivating actors to accept sanctions.

So, first, why should the sanctioner be motivated to undertake the cost in sanctioning non-cooperative behaviour? Second, why should the sanctioned actor accept the rebuke? I will argue that neither question can be answered without assuming some normative motivation in actors. Let us first focus on the second question. To do so, let us quote an important passage from Coleman at length:

> …the sanctioner may paradoxically have depended on some implicit support from the person being sanctioned, that is, the sanctioner may have felt that the person accepted the normative definition of what action is right and recognized that the action carried out was wrong. Second, the sanctioner in either case may have been able to bring up the event in subsequent discussion with others who shared the same opinion or feeling about the event and would provide encouraging comments in support of the disciplining that the sanctioner carried out.
> …whether the sanctioner depended on implicit support from the target actor or on subsequent approval from a third actor, there was an assumption concerning what is right. That is, both mechanisms on which the sanctioner may have depended for support are based on a norm defining what is the right action…or what is the wrong action. The norm, prescribing what is right or proscribing what is wrong, gives a sanctioner some presumption that his action will elicit approval from those who hold the norm. Thus the existence of a norm provides for a potential sanctioner some expectation of receiving approval from the holders of the norm. (FST: 283).

I will make four claims regarding Coleman's argument in this section, all of which I shall show to be in contrast to the assumption of self-interest.

First and most simply, we may question what is meant by the "paradoxical" assumption of the sanctioned actor partly accepting the definition of what is right. For, if the actor who failed to act in accordance to the norm also *accepted* the definition of what is right, why did he not act on that normative motivation in the first place? This would seem to be perhaps explained by weakness of will, or myopia. But no such assumptions about motivational conflict which allows an actor to act against what he judges to be best is made. While Coleman admits the existence of multiple interests, this does not explain why the conflicting interests are resolved in favour of normative versus selfish interests by sanctioning.

The second point concerns the assumption that an actor can be sanctioned and accept the sanctions *if he also assumes the sanctioner to be self-interested*. If actors are concerned only with attaining material goods, then an expression of disapproval does not express anything normative at all. That is, an expression of disapproval for non-cooperation, where the sanctioner is only concerned with his own gain, expresses only a sentiment such as "you did not put money in my pocket by not acting for the benefit of the group". If we do assume that the actor being sanctioned holds a normative definition of what is right or wrong, then why would such a rebuke, which has no normative content whatsoever, motivate the actor? Such a rebuke from a sanctioner does no more than to chide an actor for not giving the sanctioner money. The claim made here is that a normative sanction must have some normative content, for example: "how could you break your promise?", "how could you benefit yourself at the expense of the rest? That isn't fair?", "it was dishonourable for you to not contribute", and so on. Each rebuke invokes a particular value (i.e., trustworthiness, fairness, and dishonour, respectively) in order to motivate the actor to behave as the norm prescribes. Without such a normative content to a rebuke, the idea of the sanction being a *normative* sanction is incoherent.

The third criticism concerns Coleman's assumption that actors will only undertake the cost of sanctioning because they are given approval by others for sanctioning a non-cooperator. Now assume away the second problem mentioned in the previous paragraph, that being motivated by social disapproval when it comes from egoists has no meaning and hence should be ineffective, which would then suggest that being motivated to sanction others by receiving social approval for so doing should also be ineffective. Instead let us focus on a simple but effective objection advanced by

Frank (1992). That is, if individuals are truly egoistic, but desire praise from others, then they should simply lie about sanctioning a non-cooperator to others to gain their praise, but not actually carry out the sanction. That is, an egoistic actor can get social approval for free by lying to others about sanctioning non-cooperators. The point is that it would seem that an actor should be motivated by the fact that he's done is what he believes is right, not simply by virtue of the fact that he receives approval for doing so. Indeed, my argument is that unless some actors are motivated by the fact that they are doing what they believe to be right or good, the whole system of explaining the emergence of norms collapses.

This argument can be seen as a simple logical argument, not one that explores different motivations and models their effects. There has been a great deal of work in the area of the emergence of norms, which suggests that cooperation can be achieved by purely self-interested actors given certain structural assumptions (including group size, repetition, availability of information, etc.), or cognitive limitations (myopia, bounded rationality, etc.). Instead, the argument here is in some senses more modest, in that it does not make particular assumptions about the distribution of values for the variables mentioned above, but seeks only to establish the theoretical claim that the explanatory structure advanced by the assumption of pure self-interest is logically incapable of explaining the facts about the emergence of norms through the use of social approval. Thus, this argument is against a kind of modelling assumption which suggests that either (a) norms can emerge without some normative motivation, and in particular (b) that social approval is a non-normative phenomenon.

Lastly, note that Coleman states that all actors concerned who "hold" the norm, "there was an assumption concerning what is right". That is, the norm is assumed to be held by actors in order to explain the grounds of others approving those who disapprove of non-cooperators, but it is the existence and maintenance of this norm that the theory is meant to explain. That is, Coleman has assumed that actors already have an "internal sanctioning system" or conscience, and hence must *assume* extant norms in order to *explain* the emergence and maintenance of norms.

All of the preceding four claims show a circularity in Coleman's argument. That is, Coleman assumes the normative properties of what is being approved of, and this is what the theory of norm formation is meant to explain. In contrast to Coleman, my argument is that actors approve of actions because the actions are normatively laudable in their own right (i.e., intrinsically), not because of the benefit that accrues to actors who disapprove of non-cooperators. This point is well made by Pettit:

> Reflecting on the automatic way in which we sanction one another's actions by approving and disapproving, you may well think that what the rational self-interested agent should do is take over this sanctioning in an intentional way and try to drive a harder bargain for the goods he offers or the bads he reserves. But here we confront an interesting and indeed pervasive paradox. When I elicit someone else's approval for an action, without intentional action on that person's own part, I enjoy a good which would not be in the offing were I to realize that the approval was provided intentionally, or at least was provided intentionally on grounds other than that it is deserved. The good of having someone else's esteem or gratitude for an action, even the good of just having him look on the action with pleasure, is something that that person therefore cannot intentionally use in exchange. If it is not enough for him to approve that he understands the merits or attractions of what I have done, if he approves only because he has an extraintentional reason (SWO- e.g., self-interested) for doing so, or only in part because of this, *then the approval loses its significance and value.* The point will be familiar. You cannot sell your approval any more than you can sell your friendship or love or trust (1990: 741; emphasis mine)[13].

[13] Offer (1997:454) notes that approval, what he calls "regard", which is motivated by the prospect of economic gain is merely "pseudo-regard".

The point is that if an actor does not praise an action for a quality intrinsic to the act, and only because the action has positive externalities for him, then the approval has no normative evaluative content.

Further, there is no "paradox" at all about the fact that the "sanctioner may … have depended on some implicit support from the person being sanctioned, that is, the sanctioner may have felt that the person accepted the normative definition of what action is right and recognised that the action carried out was wrong" (ibid.).[14] What Coleman refers to as a paradox is, as stated as C5 above, in fact a *prerequisite* for assuming that the act of normatively sanctioning someone can be rational. Namely, assuming some normative motivation within the actor being sanctioned is a necessary condition for effective social, normative, sanctioning to work (see C5b.1, above). In the same way that it would be a waste of time (and irrational) to chide an actor for failing to understand integral calculus if the actor's grasp of mathematics was non-existent, it would be irrational to normatively sanction actor if one did not assume that the actor implicitly understood the normative premises upon which the normative condemnation was based (see C5a.2).

Nonetheless, Coleman does offer further explanations of how norms and normative motivation come about. I will show that the proposed mechanisms in FST are insufficient for the task of explaining normative compliance, and that these mechanisms do not overcome the problems I have discussed in this section.

[14] I discuss this issue at length in an essay "Shame and Self-Control: Resolving Coleman's Paradox and Defining Self-Conscious Emotions" (Unpublished MS. Available from author upon request).

3.2 Internalization, Identification, and Legitimacy

I believe that Coleman has three strategies to explain how we uphold norms rationally or why normative beliefs exist, which are the internalization of norms, identification with the socializing actor or principal, and whether the norm in question is legitimate. The first two mechanisms of internalization and identification are interdependent, and both rely further on his conception of legitimacy.

First, what is it to internalize a norm, and why does it matter to the emergence of norms? Coleman replies with the following: "…since norms are devices for controlling actions in the interests of persons other than the actor, internal and external sanctions constitute two forms of policing: internal and external policing. The process of creating an internal policing system is part of a broader process which is ordinarily called socialization. It is the installation in the individual of something which may be called a conscience or superego; I will call it an internal sanctioning system." (FST: 294). Now, such internal sanctioning systems are necessary for social sanctions and shame. An individual must be capable of internal sanctioning (guilt) if informal external sanctioning (shaming) is to have any effect whatsoever. But why would an actor internalize any norms at all? Given that not only parents, but also nation-states, organizations, and the like, attempt to induce the internalization of norms, surely a rational actor, given common knowledge of egoism, could only suspect others of attempting to modify his interests for their own benefit. The question that a theory of "rational" internalization must addresss is why the case of socialization is any different from the case of insincere regard being used strategically for material gain? Put rather brutely: when a parent sanctions a child for committing a

wrong action, why does the child trust the parent that it is in the child's interest to change their behaviour?

Coleman's answer suggests that a process of identification must take place: "A major component of socialization is an attempt to get the individual to *identify* with the socializing agent" (FST:295; emphasis in orginal). Now, it is unclear how "identifying" with someone is somethig that one could do *purposively*. Further, Coleman offers no account of what *what* specifically, in terms of *content*, is identified with. Of course we know from basic social psychology that group identification occurs at many different levels: we can identify with a group because everyone is wearing the same colour, because everyone believes in God, because our interests are interdependent, and so on. Yet are any of these forms of identification, strictly speaking, rational? At the very least, it is *odd*, if not irrational, to identify with people who share only an interest in promoting their own selfish welfare.[15]

Of course, one must share with Coleman his worries about whether or how rational choice theory should address the issue of the internalization of norms:

> To examine the process by which norms are internalized is to enter waters that are treacherous for a theory grounded in rational choice…nevertheless..individual interests do change and individuals do internalize norms. It would be possible to ignore the latter fact and construct a theory that assumed all sanctions were externally imposed. But such a theory would be weaker, because it could not be used to predict the conditions under which and the degree to which norms would be internalized, and less correct, because predictions based on it would fail to take internalization of norms into account." (FST: 292-3).

[15] For an account in which the adoption of normative beliefs and social approval is based in a non-rational manner, see Sugden (1986: chs. 8-9).

Thus, Coleman is at least aware that rational choice theory must take the internalization of norms seriously as an explanatory necessity in the explanation of norms. I am in agreement with Coleman that we will do better to explain variance in the internalization of norms, rather than to simply rely on formal models of the emergence of norms which assume some random degree of internalization (Macy, 1993: 820), or to basically disregard the relative effect of internalization altogether, as is Hechter's strategy (1987: 62-9).

Thus the question now is why an actor identifies with one individual or collective over another. Surely it is in the interests of many to attempt to gain such identification, and the individual must be wary of attempts to be duped by others. Why would one principle or content to group identity be preferable over another? There is a potentially infinite set of possible coalitions that could be formed to promote shared material gain, so how do we select some groups over others to benefit from our actions? (In sociological parlance, why do we adopt one role over another?) Some conception of the group's norms being *legitimate*, or endorsable from the perspective of what may be considered best for all, seems to be Coleman's solution to the problem of identification.

We began this section noting Coleman's distinction between conjoint and disjoint norms; the former being those norms in which the target actors and the beneficiaries are the same, the latter where the targets and the beneficiaries are separate. This is the key to his notion of *legitimacy*:

> The supraindividual character of norms lies not merely in the fact that sanctions are sometimes imposed collectively… It also lies in the fact that rights to control a certain class of actions of an actor are regarded, by actors in the system, as held not by the actor

> but by others. To state it differently, the norm and the use of sanctions to enforce it are held to be *legitimate* by each of the beneficiaries. This legitimacy is evident not only because target actors *accept the definition of an action as wrong* and *accept* a sanction that *could be resisted*". (FST:325; emphasis mine).

Here Coleman partly seems to suggest the line of argument stated by C5-6, which is that accepting a sanction is a voluntary act, in that it helps an actor achieve aims that might otherwise have to be foregone. But the fault here is that Coleman is trying to explain how norms which develop and survive do so because of wealth maximising rational self-interest, but must presume that the legitimacy of the norms is what motivates compliance.

In suggesting that a sense of legitimacy explains why actors identify with others, and why identification explains the internalisation of norms, Coleman seems to endorse the view that approval requires normative beliefs on the part of the actors, and that this normative belief is reflected in the content of the approval or disapproval. However this assumes a missing step in the argument, wherein justification of certain standards (over others) as legitimate has already taken place.[16]

So how does this legitimacy come about? Coleman explicitly borrows from social contract theory to ground the claim that we can explain the formation of groups rationally by postulating that actors use the same devices described in the social contract tradition to create associations for mutual protection or mutual gain (FST: 330). In a nutshell, legitimacy is to be found to the degree that the targets of norms are also the beneficiaries; that is, the norm serves to promote mutual advantage (cf. FST:

[16] I assume that normative justification is the defining characteristic of legitimacy, which I discuss further below.

347, and 88, on the acceptance of authority). Thus, Coleman sees the value of procedural legitimacy as being a function of the mutual advantage secured by co-operation. The question now is the following: can an egoist actually be motivated by the mere fact of mutual self-interest? Obviously not, for this is exactly the mixed-motive situation of the prisoner's dilemma (PD), involving a clash between individual and collective rationality, which is the very problem that motivates the collective action *problem* in the first place

4. Conclusion

The theoretical point made here can be seen as drawing on Brian Barry's (1995) criticism of Gauthier's (1986) *Morals by Agreement*. Gauthier seeks to explain the emergence of morality (and hence moral *qua* social order) from the situation of self-interested actors in the PD. Hechter (PGS: 124n) notes that Gauthier's framework is bound to fail because it cannot enforce its norms without a group. Barry's criticism of Gauthier essentially shares Hechter's criticism that actors have no incentive to uphold their agreements once they have been made, but instead of arguing that the only way that we could have norms is to have an external authority vested in a group (for how could it form?), Barry suggests that individuals are simply motivated by the fact that they are acting in a way that is *mutually*, not *individually*, beneficial; that is, Barry places what he calls the "agreement motive" but for our purposes may just be called (following Rawls (1971)) the "sense of justice", directly into the utility function, which means that the problem of enforcement does not become subject to the metanorm regress problem. I quote Barry's argument at length, for it expresses that which I have tried to argue for here:

> There is…a problem about moral sanctions which does not have an analogue with any problem about legal sanctions. If we understand the positive morality of a society as 'the morality actually accepted and shared by a given social group', it is surely clear that this is a collective good in the technical sense that it provides diffuse benefits to the members of society- benefits which in the nature of the case cannot be confined to those who contribute to the supply of the good. Now, maintaining the fabric of a society's positive morality requires constant effors: people have to take the trouble to form judgements about the conduct and character of others, and then undertake the sometimes unpleasant task of communicating their conclusions to those concerned. But this creates the familiar problem of collective action…With legal sanctions, this problem can be overcome…But there is nothing analogous that can be done to overcome the collective action problem posed by moral sanctions.
> But behind the problem lies a deeper one that goes to the heart of moral sanctions themselves. Is it even intelligible that there could be such a thing as moral sanctions in a society of self-interested people, at any rate so long as they were rational? Think again of a case in which I express moral condemnation of someone for acting in a certain way. The idea is supposed to be that this will help to deter this person and others from acting wrongly in future, but how exactly does this deterrence work?…It is giving the person the specific kind of hard time that consists in beign told that what one had done is morally wrong. But what sort of hard time is that in a world of self-interested individuals?…Unless my moral condemnation cuts some ice with you, it will not cause your sense of what is in your interests to shift in the desired direction…In real life, moral sanctions work because other people are able to activate an appropriate internal response, so that criticism (if we recognise its validity) triggers off feelings of guilt or self-reproach. But why should anybody in a world of self-interested people have such feelings? (Barry, 1995: 35-6).

I argue similarly to Barry,[17] namely that actors cannot rationally create social order, or follow norms, unless their reasons for complying with legitimate obligations is partly *because* they are legitimate (which can be combined with self-interest). Further, it is the fact that persons sincerely believe in normative facts that explains how social sanctions can influence actors to act in the collective interest. Another way to put the

[17] Rawls makes a similar claim but not in the context of morality as a public good per se: "Among persons who never acted in accordance with their duty of justice except as reasons of self-interest and expediency dictated there would be no bonds of friendship and mutual trust…But it also follows from what has been said that, barring self-deception, egoists are incapable of feeling resentment and indignation. If either of two egoists deceives the other and this is found out, neither of them has a ground for complaint. They do not accept the principles of justice…nor do they experience any inhibition from guilt feelings for breaches of their duties. As we have seen, resentment and indignation are moral feelings and therefore they presuppose an explanation by reference to an acceptance of the principles of right and justice. But by hypothesis the appropriate explanations cannot be given" (1971: 488).

point is that if actors cannot already have the capacity for guilt (as a function of having normative beliefs), then no other actors could ever use shame to uphold norms.

The point, then, is that the desire to act on normative beliefs must be assumed in the utility function of actors if the rational choice account is to explain the genesis and maintenance of norms through the mechanism of social disapproval. There have been many attempts to model these aspects of utility functions and the theory of the evolution of norms through formal and informal controls would do well to incorporate such models. I thus conclude that the possibility of explaining normative sanctioning in real world groups without assuming some normative beliefs or moral values is, as it is presented by Coleman and Hechter in any case, unsuccessful. To achieve success will require developing a positive theory of normative motivation which can explain the *content* of social approval and disapproval by referring to the values involved.[18]

[18] For the best (in my view) recent attempts to develop a positive theory of normative motivation see Goldfarb and Griffith (1991), Rabin (1995) Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Boudon (2001).

Bibliography

Barry, B. 1995. *Justice as Impartiality: A Treatise on Social Justice, Volume II*. Oxford: Oxford University Press.

Bolton, G. and A. Oxenfels. 2000. 'ERC – A Theory of Equity, Reciprocity and Competition'. *American Economic Review*, 90: 166-193.

Boudon, R. 2001. *The Origin of Values: Sociology and Philosophy of Beliefs*. Piscataway, NJ: Transaction Publishers.

Buchanan, J. 1965. 'An economic theory of clubs'. *Economica*, 32: 1-14.

Coleman, J.S. 1990. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Deutsch, M. 1985. *Distributive Justice*. New Haven, CT: Yale University Press.

Elster, J. (1983). *Ulysees and the Sirens*. Cambridge: Cambridge University Press.

Fehr, E. and K. Schmidt. 'A Theory of Fairness, Competition and Cooperation'. *Quarterly Journal of Economics*, 114: 817-868.

Frank, R.H. 1992. 'Melding sociology and economics: James Coleman's *Foundations of Social Theory*'. *Journal of Economic Literature*, 30: 147-70.

Frey, Bruno. 1997. *Not Just for the Money: An Economic Theory of Personal Motivation*. London: Edward Elgar.

Frey, B.S. and R. Jegen. 2001. 'Motivation Crowding Theory: A Survey of Empirical Evidence'. *Journal of Economic Surveys*, 15: 589-681.

Gauthier, D. 1986. *Morals by Agreement*. Oxford University Press.

Geertz, C. 1966. 'The rotating credit association: A 'middle rung' in development'. In *Social Change: The Colonial Situation*, Immanuel Wallerstein (ed.). New York: John Wiley, 420-6.

Goldfarb, R.S. and W.B. Griffith. 1991. 'Amending the Economist's "Rational Egoist" Model to Include Moral Values and Norms, Part 2: Alternative Solutions'. In *Social Norms and Economic Institutions*, K.J. Koford and J.B. Miller (Eds.). Ann Arbor, MI: The University of Michigan Press: 59-84.

Hechter, M. 1987. *Principles of Group Solidarity*. Berkeley, CA: University of California Press.

Hechter, M. 1991. The emergence of cooperative social institutions". In *Social Institutions: Their Emergence, Maintenance and Effects*, Michael Hechter, Karl-Dieter Opp and Thomas Voss (eds.). New York: Aldine de Gruyter, 13-33.

Lind, E.A. and T. Tyler. (1988). *The Social Psychology of Procedural Justice*. New York: Plenum Press.

Macy, M. 1993. 'Backward-looking social control'. *American Sociological Review*, 58: 819-36.

Mansbridge, J. 'Starting with Nothing: On the Impossibility of Grounding Norms Solely in Self-Interest'. In *Economics, Values and Organization*, A. Ben-Ner and L. Putterman (eds.). Cambridge: Cambridge University Press: 151-168.

Milgram, E. 1997. *Practical Induction*. Cambridge, MA, Harvard University Press.

Miller, D. 1976. *Social Justice*. Oxford: Oxford University Press.

North, D. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Offer, A. 1997. 'Between the gift and the market: The economy of regard'. *Economic History Review*, 3: 450-76.

Pettit, P. 1990. *Virtus Normativa*: Rational Choice Perspectives. *Ethics*, 100: 725-55.

Rabin, M. 1995. Moral Preferences, Moral Constraints, and Self-Serving Bias. Unpublished MS, Department of Economics, University of California, Berkeley.

Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Schelling, T.C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Sugden, R. 1986. *The Economics of Rights, Cooperation and Welfare*. Oxford: Basil Blackwell.

Williamson, O. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.

ENDNOTES