How Social Norms Matter?
Critical Reassessment of Coleman’s Accounts*

Jaehyuck Lee
Sogang University

Three points are claimed: first, that norm can be analyzed from a purely rational stance, and that the rational reductionist strategy is a better way to understand social norm. Second, contractual perspective is emphasized such that social norm belongs to a (set of) Nash equilibrium realized by individually rational strategies. Maximization over extended time periods (at the individual level) is rendered as the key factor from which the rational character of norm (at the social level) is derived. Third, social norm inherently conveys a degree of sub-optimality because of the time consistency problem. And the sanctioning system and the normative nature of social norm will be reinterpreted in our conceptual framework. Each of these points will be discussed in detail both theoretically (by using well-defined game-theoretic concepts) and empirically (by relying on various real life examples) in subsequent sections.

Keywords: social norm, rationality, time consistency, Nash equilibrium, suboptimum

There is some irony. The concept of social norm has occupied the central place in sociological explanation of human behavior. It would be even redundant to recall that the problematization of moral-normative order in modern society was the central theme for founding fathers of sociology, especially for Durkheim (Durkheim 1993; Hunt 2002). When, on the sway of the Parsonian functionalist framework, Wrong (1961) criticized the “oversocialized conception of man in modern sociology,” he was certainly referring to the excess of focusing on norm’ influence on individuals in sociological accounts. But, as Therborn (2002: 864) observes, “While the general external perception of sociological (wo)man is a person driven by social norms, within the current mainstream of the sociological discipline it is hard to find any substantial treatment of norms and of normative action.” Just when explicit sociological discussion of social norms began to wane, other social scientists were being gradually intrigued with social norms (cf: Hechter and Opp eds. 2001). Over the recent years the theoretical

* The author is grateful to three anonymous reviewers of KJS for their detailed and helpful comments on the first draft. Direct all correspondence to Jaehyuck Lee, Department of Sociology, Sogang University, Seoul, 121-742 Korea (Phone: 82-2-705-8366; E-mail: dalpang22@paran.com).
concerns on social norm has rapidly increased among legal studies (the ‘law and society’ field),
economics (the ‘law and economics,’ and the economic growth field partly related with the
concept of social capital), and social science in general. It is especially among legal scholars
who, out of “a vision of law as but one method of social control” (Freeman 2006: 1) rekindled
their interest in the social norms; as Etzioni (2000: 157) put it, “legal scholars rediscovered
social norms” (for a review of current states, see Posner ed. 2007; Freeman ed. 2006). The
increasing perception that social norms are significantly important to legal, political, and
economic theorizing should be welcomed by the discipline of sociology (‘we already said so!’).
But the grim irony is: “Sociologists have long studied the creation, transmission, and
enforcement of norms as well as the pairing of norms with social roles. The field of sociology,
however, has not had much influence on the scholars in other disciplines who have recently
become interested in norms. Partly because sociologists themselves have been unable to
coalesce around a dominant paradigm, the alien newcomers have had difficulty finding much
worth borrowing” (Ellickson 1998: 542).

Social norms matter, and matter greatly. No one denies this; the point is rather to offer
succinct explanations on ‘how they matter exactly, and in what sense?’ We need some
consistent theoretical framework to capture the essential nature of social norm, and the features
of emergence process of social norms. In this paper we will present a theoretical rationale for
analyzing social norm in terms of the rational choice (RC) framework. The primary argument
of this paper is that social norm is a particular type of equilibrium state, where rational
constraint binds mutual incentives among straightly rational agents in their longer-term
interests. Building on James Coleman’s basic perspective and using game theoretic tools and
dynamic optimization concepts, we will expound the points that 1) social norm can be
“rationally rendered” in the word’s full sense; 2) that social norm (along with other types of
‘code of behavior’) can be characterized as a certain equilibrium state; 3) but that social norm
is fundamentally “sub”-optimal in its character due to the problem of incentive binding. We
will figure out how we can capture the essential features of social norm such as its proscriptive
nature (“normative”), its informal and self-enforcing characteristics, and its ‘second best’
nature. Out of these configurations of social norm, especially its ‘second-best’ nature, we will
discuss how social norms (i.e., non-legal norms) are distinguished from and contrasted to other
types of code of behavior such as social convention and formal law, and how the working of
social norms, many of which have been confused and puzzled among researchers, can be
plausibly and reasonably interpreted in various real life cases.

We start our discussion in the first section by claiming the advantage of the rational
reductionism of RC. We proceed to discuss on the methodological issue of linking rationality

---

1 The individual level assumption of “hyper-rational agent” is not an ontological one, but a methodological one; that is, agents and players in this paper are ‘representative agents’ contrived for the purpose of modeling analytically the working mechanism of the ‘system’ phenomenon (i.e., social norm).
and norm, and indicate three fallacious research tendencies related with it. After contrasting Coleman’s approach to Elster’s, we reconsider Coleman’s treatment of social norm in the following section, especially in the context of the Coase theorem and the Prisoner’s Dilemma game setting. Next, on the basis of contractual perspective in an extended time horizon, we provide our own definition of norm in the third section. Then, we elaborate each of our specific points according to our conception of social norm in the following section.

RATIONAL RENDERING OF NORM?: A METHODOLOGICAL CONSIDERATION

Social norm is arguably one of the most profoundly important but notoriously ill-analyzed concepts in sociology. In his seminal work, *Foundations of Social Theory*, Coleman starts his analysis of social norm by noting: “Much sociological theory takes social norm as given and proceeds to examine individual behavior or the behavior of social systems when norms exist. Yet to do this without raising at some point the question of why and how norms come into existence is to forsake the more important sociological problem in order to address the less important” (Coleman 1990a: 241; italics added). This is an exactly relevant point about where the vacancy of hitherto sociological analysis of social norm has been, and where the rational choice perspective could contribute toward a more satisfactory understanding of the various social phenomena involved with social norm. Two gingerly raised questions are, then, why it must be the rational choice paradigm over others, and, on a more basic level, how a conflation of norm and rationality could be possible. Before elaborating our analysis of social norm based on rational choice perspective, we need to provide a brief assessment of our own concerning on these two meta-theoretical questions.

The rational choice strategy is to explain the target ‘system-level phenomenon’ (or the aggregate variable) through its constituent elements (or individuals), not by bringing up other aggregate level variables. In its explanation, thus, some explicit considerations are paid on the so called ‘from-micro-to-macro link’ part, yielding much greater analytical power to its explanations (Lee 2003). The social norm phenomenon fits well as the major target of rational choice theory, as Coleman sees that “The emergence of norms is in some respect a prototypical micro-macro transition, because the process must arise from individual actions yet a norm itself is a system-level property” (Coleman 1990a: 244). When considering various and elusive appearances of norms and confusing accounts about them, we believe that sharper analytical glance is called for in our efforts to grasp their essence. Everybody apparently “knows” what it is, but each one talks about it in every other ways. Social norm is a “social” phenomenon, i.e., “faits sociaux” (social facts; Durkheim 1895); we should be able to access its thrusting objective feature out of such confusing and elusive appearances. Approaching such target as norms, the rational choice perspective certainly has an endowed edge to start. Our approach in
this paper is to characterize social norms as a particular type of equilibrium state coming out of strategic interactions. Eventually our methodological position concerning the first question will be clarified as arguments go.

Three Fallacies in Approaching the Norm

In this section, we focus mainly on the second question, on the relationship between norm and rationality. How much social norm be rendered as rational? Is it a desirable direction of approach? Related with this issue, we first raise the point that there are three undesirable tendencies in approaching the norm phenomenon, especially when we want to figure out general features of social norm; Panglossian functionalism, misplaced psychologism, and methodological constructivism. The Panglossian fallacy is maybe easier to explain than the other two. Norms are as old as society is; and a naive functionalistic approach would direct us to search first of all some good reason (Raison d’être) or benefits a certain existing norm may have either for an individual member or for the community. The basic stance of the ‘classical law-&-economics’ approach concerning any spontaneously evolved system is rather positively biased toward an efficiency view. And norms belong to that category of ‘spontaneous order’ (Hayek 1973), and the Panglossian danger is there. The ‘efficient-norm thesis’ (e.g., Ellickson 1991) or ‘efficient-law’ perspective among the mainstream ‘law and economic’ field (e.g., Posner 1992: ch.2) seems to be widely held in the economics-based orientation, along with rational choice researches. This paper will see that norms are “largely” efficient, but in a fundamentally restricted way. It is far from the Panglossian sense of efficiency.

Next, the misplaced psychologism. Certainly, norm is internalized in an individual, and works through some mental/emotional channels to propel or force an individual to act in certain ways. No one can deny that investigating these psychological or neurological aspects of norm-following behavior leads to valuable knowledge to understand the norm. But the important point is that norm is a quintessentially “social” phenomenon, norm itself belonging to the ‘system’ level category. It works through and on individual psychological channels (guilt, shame, envy, zeal etc.), sure. But pro tanto; that never means that the essential nature of norm therefore could be captured in those psychological domains too. As ‘social codes of behavior,’ norm, custom, or law all share the property of indicating some social “pattern,” not its constituting elements. And such a (social) pattern, though being composed of some elements, certainly is a different thing from its constituent elements. There is unfortunate tendency of misleading norm research into looking at this and that bits of ‘elements’ (i.e., knowledge about various psychological symptoms and emotional mechanisms) and regard them as the very ‘pattern’ itself, — norm, as a social category, vanishes.

Lastly, the constructivist error. Although this line of ‘constructivist fallacy’ is very well known in terms of Hayek’s (1973) sense, the constructivism trap we here claim has little to do with political philosophy; it is about methodology. A better way to deliver its meaning may be
to indicate the most commonly observed fallacy related with it; that is, to mistake “methodological” individualism for “ontological” individualism. When ethnographers explain such cultural practice of village has some reason for its existence, they’re not saying that most people involved in those ‘reasonable practice’ are actually think or act in that way as if a single person thinks or acts intentionally in that way. A lot of those actual persons surely more than often think in other way and “err” in other way; but all these people are involved in and (largely unintentionally) “contribute” to sustain those (‘system level’) practice. When it is allegedly known that typically men err in action and fail so often to be rational, this fact does not necessarily mean that any system property out of interactions of such erring men also tends to err or to be less rational. Further, that same fact neither necessarily mean that any explanation of those system by assuming rational men for the sake of analysis (i.e., “methodologically”) must be wrong or invalid. Those who mistakenly take the assumption of “ontological” individualism are prone to demand that any analytic explanation pertaining to the mechanism of the ‘system’ should also hold its validity in the description of every concrete cases. This fallacy thereby leads to an overly inductive type of research strategy.

**Elster’s Contrasting Stance**

Our discussion about three undesirable (or fallacious) research directions indirectly helps us to assess and evaluate Coleman’s overall position. As will be shown, Coleman’s rational choice approach to norm can be judged as clearly off from the second type of danger, and largely safe from the third one too; as with the Panglossian trap, we think his theory is more or less vulnerable, and we will soon explain why. Now on the question of conflation between rationality and norm, we need to assess carefully somewhat contrasting lines of argument from Coleman’s. Elster’s (1991, 1990a, 1990b) more “conservative” stance on this issue is well-suited to our discussion. Note that a standard RC research strategy (Coleman and Farraro 1992; Lee 2003) leads to one kind of reductionism against another; it favors rational reductionism over a psychological one. In adopting the rational choice perspective, Elster, among others, is very explicit in opposing the rational reductionism of norm: “I believe that social norms and rationality must be placed at the same level. Since I take it that nobody would deny that rationality is an individual-level phenomenon, the same must be true of social norms” (Elster 1991: 114). Elster’s opposition is understandable when we look at his conception of social norm: “A defining feature of social norms ... is that they are not outcome-oriented. ... Social norms are either unconditional or, if conditional, are not future-oriented. ... Social norms have a grip on the mind that is due to the strong emotions they can trigger” (1991: 111; italics original; also see Elster 1990b, 1989). As a natural extension of this conception, he proposes an inductive strategy: “Careful examination of the diversity of norms in different societies may provide tentative generalizations, which in turn may suggest possible causal mechanisms. Although progress will be slow, it will at least be speedier, I believe, than by proceeding from

Elster’s notions and stance on norm are objectionable on three grounds. First, note the fact that Elster locates norm at the level of individual, which, unless his explanation is confined strictly to the internalized norm, is untenable to many sociologists, let alone RC theorists, whose general conception of norm would be to regard it as a collective phenomenon, “a property of a social system, not of an actor within it” (Coleman 1990a: 241). As Malinowski well indicated earlier, social norms “are not sanctioned by a mere psychological force, but by a definite social machinery of binding force” (1932: 55). Even though guilt or shame feelings (Elster 2003) are obviously important accompanying part of social norms phenomenon, focusing on those psychological aspect as an essential feature of social norm is a quite wrong-headed approach. Second, pointing out how most of Elster’s suggested examples of “not outcome oriented” social norms can be reinterpreted as strictly outcome-oriented ones from “the rational pursuit of personal advantage” (Lukes 1991: 145), Lukes made very plausible arguments that “norms are not to be contrasted with outcome-oriented motivation. ... [People] follow norms for a very wide variety of motives, many of which can quite naturally be seen as oriented to the promotion of some future good” (148; italics added). We do not believe that people are just blindly caught in the normative “grips” of norm; consciously or symbolically (Posner 1998), they take the norm through an active process of interpretation, selection, adhesion and rejection in terms of their specific social situations. Elster’s explanation of social norm as a kind of psychological behavior-motive, therefore, is in most cases either unsatisfactorily incomplete or methodologically sterile. Finally, and most importantly we can cast the following doubts against any research position (including Elster’s) that sustains a basic contrast between rationality and social norm: Even if it is true that internalized social norm comes into play (on an equal footing with rationality, as argued by Elster) in each individual’s action, one still should theoretically answer the questions of “why” and “how” such a norm came into existence and was shared by people at the first ground. Quite contrary to Elster’s inductive strategy, therefore, we claim that rational reduction is a better way to understand social norm, which point is nicely put by Lukes; “If the task at hand is to explain why people follow norms and why they follow these norms rather than those, the most promising strategy will be one which will make as much norm-following as possible come out as rational” (1991: 147; italics added).

REASSESSMENT OF COLEMAN’S NORM THEORY

Coleman’s (1990a, 1990b) conception of norm is that (1) it belongs to the system level phenomena, (2) it is purely outcome-oriented, and (3) it can be analyzed primarily by the rational reductionist strategy. The power of Coleman’s analysis derives mainly from its
simplicity and consistency. One could not find much to dispute — unless on the meta-
theoretical level — in Elster’s cautious and case-by-case interpretations of various social
norms; some of them may be rendered trivial, but they are at least not false. In contrast,
Coleman’s construction of social norms is simple and consistent enough to be easily falsifiable
by potential counterexamples, or differently put, for one to easily disagree with Coleman’s
interpretation of particular examples of social norm. In his retrospective evaluation of
Coleman’s norm theory, Elster (2003: 303) while advocating an emotion as a key aspect of
social norm, remarks that “To conclude, I believe that Coleman focuses too much on
efficiency-promoting social norms and that even for theses norms he fails to show how their
promotion of efficiency helps to explain them.” In this paper we will figure out how norms can
be efficient (or inefficient), and in what manner. This paper starts with Coleman’s framework,
and tries to refine it; and in this sense this paper can be regarded as a Colemanian reply to
Elster’s above allegation of “failure”.

Coleman’s Account in the Context of the Coase Theorem
Coleman’s account begins with the assertion that the demand for a norm arises due to the
behavioral externalities: “The condition under which interests in a norm, and thus demands for
a norm, arise is that an action has similar externalities for a set of others, yet markets in rights
of control of the action cannot easily be established, and no single actor can profitably engage
in an exchange to gain rights of control” (1990a: 251). Social norms help internalize such
externalities (by implicitly adjusting the ‘rights of control of actions’), thus bringing overall
efficiency gain to the group. We claim that such Coleman’s conception of social norm can be
better presented in the context of the Coase theorem (Coase 1960). The Coase theorem, or a
particular version of it (‘weak version’), says that with zero or negligible transaction costs,
social optimum would be achieved by internalizing any externalities through free bargaining
between concerned actors, irrespective of the initial distribution of property right.

Coase Theorem: Consider a bilateral contracting situation in which the parties are
rational with respect to their individual self-interests, and in which the parties can agree
on any contract without incurring transaction costs. Then the allocation of real goods
after contracting will maximize total welfare regardless of the initial allocation of real
goods. [Hermalin et al. 2007: 24]

There have been various attempts to refute or refine the theorem itself as well as its
implications (e.g., Demsetz 1967; Cooter 1982; Coleman 1990a: ch. 10). But generally
speaking, the Coase theorem, as a particular version of partial equilibrium theory, is closer to a
tautology with respect to its vagueness of substantive content, and it is this vagueness that
renders the theorem hardly refutable.² In a nutshell, Coleman’s norm can be interpreted as an
**implicit pricing system** for the actions (of “target” agents) which generate some externalities (for the “beneficiary” agents), which is backed up by a proper sanctioning mechanism. Note when and why non-legal norm, rather than formal law or market, arises; it is generated from the situation where there is a lack of either proper *market pricing* system or legal property rights for those externalities. Thus, social norm is construed as a system socially devised for curing (economic) market failure, if we view society as a kind of a ‘grand social market’ of which economic market is just one component (Arrow 1971).

This interpretation of norm as an implicit pricing system can be nicely illustrated by the example of the norm against smoking. First, it is clear that smoking has a negative externality to nonsmokers (cf: see Opp [2002] about smoking norm case). Until a decade ago, smokers did not feel much pressure to conform to the norm against smoking in public places because, for whatever reason, it was believed and implicitly shared by most people that the right to smoke belongs to the smoker. Suppose, again for whatever reasons (possibly due to recent scientific evidence on the potential harms of secondary smoke to the nonsmokers), there has been a change in this implicit right to smoke privileging nonsmokers. As a result, smokers now must pay some additional costs for smoking — for example, psychological pressure or the risk of unintentionally provoking nonsmokers — imposed by the social norm against smoking, over and above the market price of a cigarette that the smokers have already paid. The ‘shadow price’ of a cigarette to smokers now becomes higher than the tagged market price. This aspect of social norm is aptly captured in Sunstein (1996)’s statement that “we can understand a norm — with respect to choice — as a subsidy or tax (939): “From the standpoint of an individual agent, norms provide a part of the background against which costs and benefits are assessed; more specifically, they help identify some of the costs and benefits of action. From the standpoint of the individual agent, this is hardly irrational, and it is hardly inconsistent with self-interest” (Sunstein 1996: 935).

Two things should be noted from this small illustration. The first one is the question about why this social norm was generated and how it could vanish or be replaced by other substitute social mechanisms. It was when there was no explicit formal articulation placing a legal or formal right of smoking into the hands of nonsmokers that a social norm was generated to

---

2 The zero transaction cost assumption plays an important role in this. Depending on how one regards the range of the component, “transaction costs,” this theorem is not very informative about the actual exchange processes. For example, private bargaining is crucial to deriving the self-correcting efficiency in Coase theorem; but in many real situations, the potential bargaining costs prevent two concerned parties from achieving a mutually beneficial agreement. However the problem lies in locating the source of this preventative cost, not just a balancing of cash accounts. If agents are perfectly honorable in revealing their intentions, and if the once-agreed-on rule is strictly enforced, then the configuration of transaction cost is easy and the result of no bargaining, after such configurations, could be still conceived as Pareto optimal. Thus, the real source of difficulty lies rather in the incentive problem and the enforceability problem of the agreed rule.

3 Demsetz in his classical paper (1967) claimed that property rights are generated to internalize externalities in individuals’ transactions. His efficiency view of property rights, which has largely affected further development in ‘law and economics’, is based on the reduction of various transaction costs.
internalize the externality. Note, then, the recent trend that has witnessed the gradual disappearance of the social norm against smoking as more formal rules such as legal ordinances or organizational rules to ban smoking in public places are being introduced in a more explicit way. Also imagine how such a social norm against smoking would vanish if nonsmokers could in some way charge an explicit amount of money for each unit of smoking. It is the deficiency of market pricing for the external effects or the lack of explicit legal right, other than the externality itself, that provides the necessary ground for the generation of social norms.

The second point is that social norm would be better understood upon the ‘efficiency’ ground in its overall sense (The reason why we put this caveat on the term ‘efficiency’ will be clear later). Note that the above example, and Coleman’s analysis in general, can well fit into the Coase theorem *mutatis mutandis*; that is, social norm internalizes the externalities, thereby leading to the social optimum state out of individuals’ interactions. Roughly stated, if every social action could be priced in some way in the so called ‘grand social market’, and if the social pressure of norm were construed as one of these *implicit* prices (sometimes additional to market prices) for the concerned actions, then according to the First welfare theorem this decentralized pricing system would lead to the social (i.e., Pareto) optimum. So, in Coleman’s framework, most social norms could be generated and maintained in a very rational fashion by individual members if, or (to a lesser extent) primarily because, it helps them achieve the state of social optimum.

Interestingly, and unfortunately, Coleman’s own assessment about the implication of the Coase theorem on his framework is somewhat different from that presented here:

The implication of [Coase] theorem for disjoint norms is that if the externalities imposed by the activity are sufficiently great that the persons experiencing them are motivated to induce the actor to cease, and if those persons have the resources to make that motivation effective, then they will do so, *whether or not a norm is in existence* (that is, whether or not the informal rights to control the action are held by them). This implies that a norm (or a law) is superfluous and would make the existence of norms or laws inexplicable.

How is the logic of Coase’s argument consistent with the existence of norms and laws? The answer lies in the assumption on which the Coase theorem is based — that there are no transaction costs in inducing the actor to cease carrying out the action when he has the rights. There are, however, transaction costs in real social systems, and *a norm can reduce the transactions necessary to achieve socially efficient outcomes*. [1990a: 261; italics added].

This is more or less a standard interpretation of the Coase theorem, and, within it,
Coleman’s question and his own answer are both valid. But Coleman’s assessment, especially in its reliance on the assumption of transaction cost, is only partly true and does not capture the whole story. The empirical validity — itself questionable, though — of the assumption of ‘no transaction cost’ is rather secondary to the issue of social norm (see footnote 2 above); contrary to widely held opinions of many researchers about ‘efficient-norm’ thesis, including Coleman, this paper claims social norm has much more things to do with than just the reduction of transaction costs. If transaction cost reduction is the main reason for sharing a social norm, then, we justifiably wonder why norm, understood as such a behavioral coordination scheme, has such a strong degree of normative grip on behavior; and why not people resort to any other less costly or more explicitly coordinate mechanism such as convention, custom, organizational rules, or procedural law; and, if the issue is concerned more with restitution than transaction cost, why not resorting to a more effective legal system (cf. “norm management”; Sustein 1996). This functionalist-type of interpretation dilutes the significant nature of social norm and becomes an easy target for various criticism (e.g., Elster 1996, 1991, 1990a; also ‘legal centrism’ tradition in Cotterrell [1998]). Depending on the degree to which we assumedly expand the boundary of the set of “market” and “prices (of behavior),” however, we could readily resolve the seeming contrast raised by Coleman between his framework and the Coase theorem. We shall raise a similar question about the implication of the Coase theorem in a different context, and discuss in detail later. The point here is that regardless of the transaction cost assumption, Coleman’s framework can still be interpreted in the context of Coase’s theorem. In this regard, Arrow’s observation interestingly coincides with the reinterpreted Coleman’s conception of norm: “I suggest as one possible interpretation that [norms of social behavior] are reactions of society to compensate for market failure. ... There is a whole set of customs and norms which might be similarly interpreted as agreements to improve the efficiency of the economic system (in the broad sense of satisfaction of individual values) by providing commodities to which the price system is inapplicable” (Arrow 1971: 22; italics added). Also interestingly, Ellickson (1991)’s work, which set the benchmark of the revival of the ‘law and society’ discipline in legal studies, reaches a very similar conception of social norms as the one implied by Coleman’s framework. After conducting careful field observations about various norms concerning the ranching system in Shasta county, California, Ellickson draws a conclusion that “in uncovering the various Shasta County norms, I was struck that they seemed consistently utilitarian. Each appeared likely to enhance the aggregate welfare of rural residents. This inductive observation, coupled with supportive data from elsewhere, inspired the hypothesis that members of a close-knit group develop and maintain norms whose content serves to maximize the aggregate welfare that members obtain in their

---

4 On this aspect of transaction cost reduction, Coleman, in a personal conversation, offered the following additional point: “Here I should have said more; the norm reduces transaction costs by a reallocation of informal rights which makes transaction (if target actors accept the reallocation) in the form of sanctions unnecessary.”
workaday affairs with one another” (1990a: italics original). However, in singling out the reduction of transaction costs as a main function performed by social norm, Ellickson’s approach is closer to Coleman’s own interpretation than are Arrow’s and ours.

SOCIAL NORM AND INCENTIVE BINDING; REFINING OVER COLEMAN’S DEFECTS

One notable characteristic of Coleman’s conception of norm, which is otherwise generally interpretable on an efficiency ground as shown above, is that he makes a clear-cut distinction between the “demand” for norm and the “effective realization” of norm. For the latter, an effective social sanctioning system and, as its requisite, preexisting social networks operative among concerned members are the key and indispensable factors: “When the social optimum is that an action not take place, the action will be constrained and the social optimum achieved only if those experiencing the externalities have the requisite social relationship to both bring a potentially effective norm into existence and make it effective” (1990a: 262). Both “time closure” (the degree of repetitiveness of a relationship) and “social closure” (the density of the network of relationships) must be high enough to overcome the higher-order public good problem of sanctioning for the realization of norm (since sanctioning itself is a ‘public good’).

A notable example given by Coleman is how the milling process of “gossip,” a typical informal sanction, could effectively sustain the community norm within a situation of high social closure (Coleman 1990a: 284-8). He did not discuss much about the “time closure” aspect, but this concern is primary in some other approaches which typically adopt repeated game theory as a tool to explain the generation of social norm. One of our main concerns is with this time dynamics which will be discussed in detail in the following section.

One problem with Coleman’s framework, and commonly shared by many game-theoretic approaches, is that it remains unexplained how those observed high closures, either social or time or both, which could potentially solve the problem of the provision of sanctioning, came into being and are maintained. Invoking high closures as a given in norm-generating situations seems another trap of losing explanatory power — it certainly works, but it “takes too much preexisting social structure for granted in explanations” (Hechter 1990: 245). We need to endogenize those components of closure in our model to overcome this kind of problem. Roughly, the idea is that people faced with a dynamic optimization problem invest in a personal relationship in rational anticipation of adjusting both components of closure. We think this is a more correct imagery of a norm situation; people are not just constrained by social structure, but also they actively react to and reproduce social structure itself.

---

5 For the classical game theoretic modeling, see Voss (2001); Taylor (1987); Sugden (1989); Kreps et al. (1982). And for an evolutionary game setting see Mahoney & Sanchirico (2001); Bendor & Swistak (2001); Axelrod (1984).
Game Theoretic Interpretation of Social Norm

Apart from the above problem, in general Coleman’s explanation of norm relies too much on the possibility of sanctioning. It is true that one main characteristics of social norm lies in its normative enforceability (in addition to pure coordination), and every social norm must be backed by some sanctioning system; otherwise the same code of behavior, called “norm,” would be transformed into less normative social convention or rituals. Compared with norms, “social convention” is a realization of one particular equilibrium point out of possible multiple equilibria over which agents have no (or a less degree of) conflict of interests in a pure coordination game-setting (Picket 1997; Sugden 1989 etc.). For example, the ‘traffic signal’ system in which we are indifferent over whether “red” or “blue” or whatever color indicates “stop,” as long as others follow the same signal. Thus, social convention is regarded in this paper rather as a pure coordination equilibrium with no (or a less degree of) inherently normative character. But, along Coleman’s approach to separate the issue of ‘provision of sanctions’ from the ‘demand’ for norm, analytically we now face the case of infinite regression of the problem of how to provide the sanctioning system which, as indicated above, itself invokes a higher-order public good problem. Norms are in flux; and it would be realistically implausible to argue that the norm generating process, distinct from the ‘demand’ for norm itself, is complicatedly demanding to the degree to which it necessarily involves all those higher-order N-person collective good problems. At this point, we need to reconsider what is meant by “sanction” in the norm generating situation. We claim that the problem of provision of sanction is only a subpart of the more general problem of norm generating process. Coleman’s flaw of exaggerating the importance of sanction in constructing norm results from his treatment of the sanction as an independent entity separate from the (demand for) social norm itself. As in the case of transaction cost in norm generation, the sanctioning system in itself is a secondary feature in defining norms.

To expound on this point, let us use the simple illustration of the Prisoner’s Dilemma game. With the usual payoff structure of the PD game and two strategies of \( D \) = defection and \( C \) = cooperation, we can easily see that the unique Nash equilibrium \( \{ D, D \} \) (in an one off game) does not lead to the socially optimal state (Pareto optimum) which corresponds to strategies \( \{ C, C \} \). The problem, in this context of the PD game, is, then, how to achieve the socially desirable outcome (social optimum from mutual cooperation) from individually rational strategies. When the game is repeated, either infinitely or finitely with an uncertain end point, the situation drastically changes. Roughly put, the well known Folk theorem states that players

---

6 As Posner (1996) indicates, it is important to distinguish ‘norm’ from ‘convention’ for clear understanding of norm’s action. The sharp distinction between norm and social convention, however, might be blurred by extending the definition of “convention.” One may regard social norm as a variation of this broad conception of social convention, such as in Young (2008) and Sugden (1987).

7 In this respect, Bernheim’s (1994) formal modeling is suggestive; one of his arguments is that the degree of “harshness” of sanctioning in a society is produced endogenously.
can achieve as a Nash equilibrium any feasible, individually rational payoff (in terms of “discounted total” or “average discounted payoff” over the whole ground of play) above the minimum level (payoff of repeating \(\{D, D\}\)) if players are sufficiently patient. Note here that (1) mutual cooperation is a self-enforcing agreement, that is, a Nash equilibrium, (2) the degree to which players take stakes in the future critically matters (high patience), and (3) there are multiple equilibria each of which depends on a specific combination of strategies. Let us put these features of the (repeated) PD game into Coleman’s framework. First, one’s individually rational strategy in each ‘stage game’ — i.e., the dominant strategy of playing \(D\) — has a (negative) externality on the other’s welfare since the other would be unilaterally worse off by one’s playing \(D\), and vice versa. Second, the social optimum in this game situation is defined as the payoff from mutual cooperation. Third, tentatively stated, social norm in this particular game corresponds to the self-enforcing agreement on the set of strategies (playing \(C\) at each stage game) which expectedly achieves the social optimum as an individually rational outcome. Then, where is a “sanction” in this story? A sanction in this game consists of any kind of, or combination of, credible threats to punish potential deviation from the agreed-upon path of play (i.e., from the “norm” in our context); and in this particular game, the punishment is a playing \(D\) — maybe only once (Tit-for-Tat), or twice, ..., or forever, the number of punishing rounds of which does not matter in an ex ante sense as far as it reduces the deviant’s payoff in future rounds enough to effectively deter the deviation. The binding force here is self-interest and the rule of reciprocity (Gouldner 1960) rather than any externally-set sanction system. In this regard, Malinowski’s observation in the context of primitive law is remarkable: “The binding forces of Melanesian civil law are to be found in the concatenation of the obligations, in the fact that they are arranged into chains of mutual services, a give and take extending over long periods of time and covering wide aspects of interest and activity. ... Thus the binding force of these rules is due to the natural mental trend of self-interest” (1926: 67). All possible sanctions are already embedded in the payoff matrix of a game; and the Nash, if achieved as a subgame perfect equilibrium, and the resulting norm should have reflected those sanctions as a part of consideration. The point here is that the problem of provision of sanction is inseparable from (and only part of) the demand for norm itself — and this point is easily extended for other kinds of game. We need to seek some other way to define the norm generating situation than through the sanctioning system.

8 One related point is that, using Hirshman’s terminology (1970), Coleman’s examples of sanction mostly take the form of “voice” only, whereas it would more plausible to regard both “voice” and “exit” together as forming a sanction. For example, one of the most punitive sanctions in primitive society was social ostracism, i.e., the termination of all social relationships with the condemned agent for good (Malinowski 1926).

9 Subgame perfect equilibrium is one kind of refined Nash solution concept for an iterated game situation. It says that a particular set of Nash equilibrium strategies is subgame perfect for an extensive form game if that strategy set gives a Nash equilibrium in every “proper subgame” of the game. See Selten (1975).
Contractual Aspect of Social Norm

Before identifying the “other way” we would like to make some considerations on the implications of the iterated PD game. As indicated above, the Folk theorem derives three main implications; self-enforcing agreement, a stake in the future, and multiple equilibria. First, let us think about what differentiates social norm from the law, which is incredibly normative. One basic distinction, other than the strength of enforceability, lies in the fact that the former is implemented in a decentralized way (Cooter 1997), whereas the latter in a centralized way, especially through the state. From this self-enforcing agreement is evident the contractual nature of a social norm (cf. Sugden, 1990; Binmore 2005, 1994). A question then arises consequently; why not make the same code of behaviors, so called “norm,” into a law? Or in a different setting, as Ellickson (1991) asks, why do people want to resort more often than not to informal norms rather than to formal laws covering the same set of behaviors for a settlement of conflicts among neighbors? One possible answer is that norms are more flexible and can be adjusted locally, both across actors and over time, to bridge the gap between the currently practical need of some code of behavior and other more general principles of behavior such as an ethical code or constitution. For example, norm, as opposed to law, is less constrained by moral ethos or general ideology — can one conceive of a law reflecting the majority’s taste of racial discrimination against the minority, when such discrimination is implicitly practiced even under the shadow of the official ideology of equal human rights? We can pursue further the reasons why norm rather than law is resorted to in some conflict settlements.

Various forms of informal sanction (gossip, public condemnation, ostracism, etc.) have one common underlying factor — restriction or extinction of continued social relationship. Ellickson (1991) found, not surprisingly to our common intuition, that Shasta County members are more likely to appeal to the law or formal insurance policy when the conflict involves strangers (typically, in his study, traffic collisions on local highway involving commuters and cattle), and more likely to informal norms (of “live-and-let-live” in his term) when conflicts are among neighbors (trespass of cattle into a neighborhood farm or building a boundary fence). There exist laws concerning the cattle trespass issues, but they were never sought after; also, the disputants could have arranged a formal contract about who bears the fence work or how to exactly divide the fence costs, but they never did so. What motives drove them to avoid the likely more effective formal resolutions and contract? Conflict among group members is a nuisance to the continuation of a profitable relationship and thus should be resolved with a minimum “scar.” Macaulay cited one businessman on the effect of continued relationship who noted that “you can settle any dispute if you keep the lawyers and accountants out of it. They just do not understand the give-and-take needed in business” (1963: 61). In terms of

---

10 This demarcation issue actually covers the whole sub-disciplines of socio-legal studies; see articles in recently edited collections of Posner ed. (2007), Freeman ed. (2006), Banakar and Travers eds. (2002) etc.
maximizing the stream of future benefits from ongoing relationships, it would be myopic to sacrifice such future benefits for a minor amount of current gain. The normative grip of norm rests fundamentally upon the threat that this continued relationship will be terminated (cf. Horne 2007, 2004). Note that this point is diametrically opposed to Elster’s claim; in Elster’s (1989) diagnosis, people are bounded by norm because they are “backward looking,” recoiled from the past socialization; in our story, people follow norm because they are “forward looking,” calculating over future consequences. The cooperative result of the Folk theorem is possible only if people have a high stake in the future. And the future stake is none other than the expectation of continued relationships among members in the future. When the future stream of benefits from continued relationships are counted, the possible current loss from imprecise resolution of conflicts among neighbors is relatively minor; and more important, pursuing too strenuously a precise resolution of each incident would be seen as “too stringent,” rendering that pursuant less desirable as a neighbor with whom mutually beneficial social relationships can be sustained. No Orange County members risked their social relationship with each other by pursuing explicit resolution through formal methods (as one county man says, “Being good neighbors means no law suits.”); and they had good and very rational reasons not to do so.

Finally we turn to the issue of multiple equilibria. As mentioned above, not all of these subgame perfect equilibria are Pareto optimum. No satisfactory game model has yet arrived at providing an analytic solution for the condition under which the Pareto optimal outcome is selected from multiple equilibria. Even with possible friction from bargaining and inequality of power, according to one speculative view, certain social institutions or social norm might help people to approach approximately to this Pareto optimal state. But the question of “how” still remains. The idea of investment in relationship (“relation capital”) implies that people endogenize the “future stake” itself in their current and expected decision making on how to behave. Conscious control over future relationships can enlarge the degree of freedom with which we can model whether, or how, the norm of cooperation can lead social members to achieve the Pareto optimum.

**SOCIAL NORM AS ‘SECOND BEST’ SOLUTION**

In this section we formulate our conception of social norm. First, as we indicate so far in elaborating Coleman’s frame, norm is essentially a rational outcome and can be considered primarily on the ground of efficiency. Unlike Coleman, however, we place greater emphasis on

---

11 But see Gintis (2000); and Fudenberg and Levine (1998) for some preliminary results about this direction.

12 Evolutionary game and the concept of Evolutionary Stable Strategy (ESS) shed a useful light on this issue. See also related discussions in Young (2008), Picker (2001), Bendor & Swistak (2001).
the contractual (i.e., mutual incentive binding) nature of social norm over the components of externality and sanction. Now we claim that the nature of social norm is better rendered from the perspective of dynamic optimization process in general and the time consistency problem in particular. The demarcation problem of social norm vs. other principles of behavior such as law (viewed as more normative) and social convention and custom (viewed as less normative but more coordinative) also can be better understood in the same perspectives. Time (or future stake) and incentive are the key decoding factors in our rational understanding of social norm.

Credible Threat and “Normative” Grip of Social Norm

Let us start with an explicit definition of social norm: Social norm is a (connected) set of Nash equilibria with the (ex ante) second-best optimum. These terms will be subsequently explained in the discussion to follow. For this purpose, we will juxtapose theoretical claims with corresponding real life illustrations. First, as mentioned above, we regard social norm as generated by any situation where actors have mutual interest in each other’s action; and this can be a situation either where some actions have external effects which are not properly compensated by the preexisting market or explicit property right; or a situation where each other’s actions must be adjusted to achieve the potential mutual benefit from such interaction. Note that the latter case does not necessarily involve externality arguments — it is rather closer to an active coordination process of actions only for the sake of forthcoming mutual profit. In either case, members face the potential opportunity for improving their respective welfare ex ante by making an implicit contract on individually rational, self-enforcing rules on their actions. The picture we visualize here about the norm generating process is similar to that of the constitution-making process, with the exception that the rules are implicit and their implementation is decentralized and local. Here, each group member, in anticipation of uncertain future events, is assumed to choose some optimal time path of behavior to maximize his stream of interests over time. The choice should incorporate a set of strategic reactions from others and the expected costs of sanctions — either the costs of imposing sanctioning upon others for their deviations or the cost enforced on the ego by others for his detected deviation. There is an initial distribution of resources and a preexisting social structure among members; each agent faces different constraints depending upon his initial resources and his social position in the group. His choice of an optimal time path should also include an optimal investment decision with regard to his relation capital.

Based on mutual agreement, if possibly set, some equilibrium outcome would be generated such that given their initial conditions and expectations, nobody would have an incentive to change his planned path of behavior (Nash equilibrium); and this Nash equilibrium should reflect the expected, socially optimal outcome streams (i.e., it must be an ex ante optimum). For example, the “norm of revenge,” Elster’s (1990b) favorite example of a “nonconsequential” norm, can be construed as a Nash equilibrium in that society for the same
reason that the strategy of threatening to “play D forever,” or Tit-for-Tat, is part of the Nash strategies in the repeated PD game; it can effectively deter the grievances perpetuated by one clan or villagers against the other if the threat were conceived as credible by both parties. And this norm of revenge is ex ante socially optimal by reducing the extent of potential mischief intentionally or carelessly done by either side. If this is so only in ex ante sense, it may be asked, then why actually execute a revenge once deviation has been detected, which certainly harm the revenger himself? This may bring to mind Elster’s (1990b) argument that revenge is rather a blind and irrational behavior, mostly done out of emotion rather than reason. But note that the above Nash equilibrium, which incorporate a revenge as a punishing strategy, is possible only if the threat of punishment is a credible reaction to an off-the-equilibrium path. Thus, the revenge should be executed even at the cost of inflicted self-harm to render the original equilibrium “subgame perfect”; otherwise, the threat would become noncredible, and as a result the original equilibrium state would break down. Logically, once equilibrium state were reached as subgame perfect, those off-the-equilibrium behaviors should not have occurred by definition of subgame perfection; revenge, if credible, would be rarely observed if the state of affairs is in Nash. Sometimes, but not frequently, we observe such revenge or intergroup feuds occurring in actual situation (or conversely, we often observe revenge not occurring when expected to occur). Does this mean that the observed revenge was not the part of equilibrium, or that the supposed Nash equilibrium state had been wrongly inferred by the observer? Many real life situations, unlike those arising from mathematical derivations, involve inherent variation and human error (“Trembling hand”; Selten 1975), and our model must make allowances for a certain number of counter-cases if tolerable.

We will soon discuss why norm following behavior, even though being a Nash with ex ante social optimum, is a fundamentally constrained solution due to the time consistency and incentive problem. Beforehand, we continue making some preliminary claims. Returning to the picture above, we claim that those Pareto-improving moves in PD game situation, unlike the pure coordination case of social convention, are necessarily normative and, as a result, not the “first best” outcome for the following reasons. Using the framework of the PD game above, some theorists descriptively and indiscriminately interpret the Nash result in the repeated game situation as ‘a norm of cooperation’ (e.g., Bicchieri 1990). Two problems arise. First, if the term “norm” is used in such a descriptive way to indicate a cooperative Nash outcome, then, we cannot distinguish norm from other pure coordination equilibria such as convention or ritual. In this PD game, the “normative” aspect of norm is incidentally imbibed into the Nash equilibrium due to the particular context of that game. In other types of game, especially those of pure coordination, however, we would not generally call the resulting Nash outcome a “social norm”. Second, if norm as a Nash set is the “first best” (i.e., incentive binding)

---

13 A classical anthropological case of feud and its resolution can be found in Evans-Pritchard (1940).
outcome, which is implied to some degree by Coleman, then where is the normatively constraining role of norm? In other words, if people, even with their considerations of the cost of sanction, agreed on a certain code of behaviors for their mutual benefit, and if those expected results were a first best in an ex ante sense, then they would voluntarily achieve the result through pure incentive with or without the norm (if the term “norm” is construed as something more than a descriptive label of any Nash equilibrium outcome). Here we need to scrutinize more carefully the meaning and implication of the words “ex ante” and “second best” in our definition of norm.

**Time Consistency Problem and Contracting on Social Norm**

Recall that we take the contractual position where the norm generating process is viewed essentially as one of making self-enforcing agreements among group members for their future conduct. The resulting Nash outcome should, then, be an optimum necessarily and only in the ex ante sense. Unless the world is assumed as one of certainty and perfect information, these implicitly agreed-on rules must reflect an inherent gap between the ex ante anticipation and the ex post realization of events. This gap means that group members always have some degree of incentive to deviate from the previously agreed-on path of behavior. Note that these ex post incentives to deviate are not the same as those at the initial stage where the Nash set has already incorporated all potential deviations and corresponding sanctions for calculating total payoffs in the ex ante sense. For illustration, let us assume there are four time periods, time 0, 1, 2, 3. At time 0, under the anticipation of all possible states to come in the future, people agree as to how to behave at time 0, 1, 2, and 3 which would maximize each and every member's expected total welfare. At time 1, in part as the result of actions carried out by some members as prescribed, and in part as the result of the good and bad luck happened differently to each member, they now find the previously agreed-on rules (made prior to the realization of states at time 1) suboptimal for time 1, 2, and 3 behavior to maximize the utilities of these remaining periods, and have incentive to amend the previous rules from the point of this new situation. The dilemma is that once they decide to make a new rule which is optimal for the expected payoffs of period 1, 2, and 3, they now have good reason to doubt the credibility of such amended agreements being honored for the remaining periods, and thus are not likely to observe the rule since they know the same demand for amending the rule set in period 1 would be made again at time 2 (and thus abate the ex ante optimality itself of the mutual agreement as of time 1). To avoid this dilemma arising from the incentive problem, they have to stick to and continue honoring the original plan made at time 0, and embrace the suboptimality of that plan.

This story is the simplest version of the, so called, “time consistency” problem, one well-known historical example of which is the patent system. The dilemma is that on the one hand, the Pareto improving move of breaking the once agreed-on rule is not time consistent (e.g., freely distributing the new invention to all members); whereas on the other hand, honoring the
time consistent rule is socially suboptimal (e.g., forcing them pay for the patent fee to the inventor). The basic implication of this problem is that in a dynamic framework, there is a fundamental conflict between time-consistent “rule” vs. *ex post* optimal “discretion”; the former has the deficiency of being suboptimal, whereas the latter has the deficiency of invoking incentive problems. Kind of “Higher discretion” over this dilemma, then, should be invoked to decide on how much “discretion” is socially optimal; or to put in other way, on how much loss from rigid rule could be tolerable to honor the general principle of *pacta sunt servanda* (promise must be kept). Norms do not exist outside of and prior to the relevant individuals. Rather, norms “must be constantly negotiated and renegotiated in the arguments and justifications surrounding specific acts” (Weiner 1976: 222).

**Second Best Optimality of Social Norm**

The above discussion on time consistency explains the difference between *ex ante* and *ex post* incentives to deviate. Note the fact that from this dynamic perspective, people have an incentive to not observe, or even to jointly change, the once agreed-on norm without any changes in their sanctioning capacity to be involved. That is, the incentive problem resulting from time consistency, not the established sanctioning system, is the most fundamental cursor of inducing the normative character of social norm as a counterpoint to such an incentive problem. This is part of the reason for claiming that sanction itself is not the primary factor in defining norm. More important, now we can see why norm is a “second best” rather than a “first best” outcome.14 As an *ex ante* agreed-on rule of conducts, a norm inherently conveys some distortion of optimality at the cost of mitigating the incentive problems that arise naturally as time goes on.

The so-called “anti rate-busting” norm in the piece rate system illustrates this point well. Given the productivity levels of workers, the employer sets the optimal piece rate, giving enough incentives to work hard. Workers, given the piece rate, want to work harder and get rewarded accordingly. But if every worker worked hard and increased productivity, then the employer would have an incentive to raise the rate optimally corresponding to the increased average productivity. With such a rate adjustment, workers, then, would earn the same wages on average as before even with their increased work efforts. Unless there is some credible commitment by the employer to not raise the rate, therefore, each worker has an incentive to constrain his own and others’ work efforts. These joint incentives to constrain average work efforts are met in conjunction with the norm against rate busting. Note that this rate-busting norm achieves the group optimum in an “*ex ante*” sense by preventing potential exploitation arising from excessive current period competition, but is a “second best” in the sense of constraining individually rational work efforts (For the employer, this second best result also

---

14 In a somewhat different context, Kreps (1992) advances a similar line of argument about cooperate culture.
holds even if he announced a prefixed rate to solve the workers’ incentive problems, since then the fixed rate would become suboptimal when workers increase their productivities.\textsuperscript{15}

Then how much distortion of ex ante optimality is in general involved in generating a norm? This is rather a matter of degree. To the extent that the distortion of optimality is conceived as tolerable by group members, social norm will be shared and maintained in a group without losing its normative force. As the gap between the ex ante and the ex post optima increases, possibly due to some external changes like new technology or new ideology, there would arise some contentions by some members against the agreed-on rule of conduct (i.e., more deviating behavior), and the norm would start losing its normative grip because its legitimacy basically rests on the (power-weighted) consensus among group members. When the distortion of optimality becomes intolerable, the once agreed-on norm will either disappear or be transformed into a nonnormative convention or ritual; and another type of norm or other principle such as formal organization or law will replace it as the new code of behavior.

For illustration, take the example of the famous Golden rule, “Do unto others whatever you want them to do unto you” (Matthew 7: 12). Two questions should be considered: First, “how this kind of norm evolves” and second, then, “why not in other forms of rule such as a law.” As to the first, we need to consider how the golden rule can lead to an (ex ante) optimum among players. The golden rule is a more active version of the principle of reciprocity; it encourages creating informal credits to others in the tacit expectation of being repaid by the equivalent. By doing so, it also embodies in part the effect of signaling (of cooperation) so as to resolve the possible strategic uncertainties among players (see Posner 2000; Ostrom 2000). Note carefully the conditional part of this rule, “whatever you want them to do unto you.” The selfish and rational, as we assume, player can easily calculate what would be the best for him or how valuable that would be if he were in the concerned other’s position of need. This means that he can also calculate the relative shadow price of his action or his resources at the current time point in comparison with the one (usually discounted by some “interest” rate) in the future or the conjectured time of such need. Usually, the same action or resource becomes more valuable in times of need or desperation. Considering this, each actor has a great incentive, especially where there are no well-organized formal financial institutions, to shift his resources and his efforts (for a particular action) over to these possible states to optimize resource allocation over time so that either some arbitrage profits are obtained from the network of relationships (“structural hole”; Burt 1993) and the variation in his overall utilities is reduced by such a smoothing. If each and every member could be convinced of being reciprocated by the equivalent (but with a different shadow price at the time of being reciprocated), all would be

\textsuperscript{15} See Akerlof’s (1974) “rat-race” story for a similar result of suboptimality of separating equilibrium. Very interestingly Posner (2000) develops a signaling equilibrium interpretation of social norm. And since every signaling equilibrium is suboptimal, it is significant to note that his conception of social norm shares our model with respect to the fundamental second best characteristics of social norm.
better off by engaging in such an informal credit system. The golden rule reflects this subtle but precise process of pooling and redistributing individual resources and efforts for the joint benefit of all members.\(^\text{16}\) Thus, if the golden rule can be sustained as a Nash strategy, which is quite possible depending on the degree of time and social closures of the group, then it could provide an implicit base to facilitate these Pareto-improving moves by group members. But note that the social optimum achieved by this norm is construed thus only in an \textit{ex ante} sense; each member runs the risk of failing to get back his extended credit later — actually, there were no “written” contracts for balancing the “accounts.” We need to appreciate more carefully one subtle point here.

According to the definition of subgame perfect equilibrium, the golden rule norm must be incorporating all potential incentives to deviate from the norm and all corresponding credible punishments at the time it is established as a Nash. That is, there should be no incentive problems \textit{ex ante}; otherwise the golden rule could have no longer been a (subgame perfect) Nash equilibrium at the first round. Once time passes and some uncertainties are resolved, however, the relative valuations on the conjectured “one’s need time” change among group members since their positions in the social structure are now different from the initial ones. Thus, necessarily the debtor’s repayments and the creditor’s expectations are not precisely matched as originally set in the previous setting. This gap between obligation and expectation, when nontrivial, abates the incentives to stick to the original equilibrium path. Again, the dilemma is, if the group members stopped honoring the original equilibrium path, then the mutually beneficial credit system would collapse, resulting in another kind of social suboptimality. By honoring the golden rule norm, therefore, each of the group members forgoes some individually rational optimum and implicitly agrees on the second best social outcome. It is in this context of time consistency that social norm, though efficient, constrains social members from achieving the first best outcome. Next, then, why must this be accomplished by a norm rather than other formal rules or institutions? Indeed, formal financial institutions have developed, and they have replaced many of this type of norm of symmetry in their implicit role of financial instrument. Also many refined tort and contract laws based on the “principle of expectation” replaced the role that social norm play in adhering social interactions. It is partially for this reason that the golden rule has become a rather vague precept instead of serving as a normative code for contemporary people. A more fundamental reason must be sought in the changed nature of contemporary social structure, that is, in the \textit{reduced importance of informal social relationship} — because of the decline of community,

\(^{16}\) In this respect, Weiner’s (1976: 222) statement about norm of reciprocity in simple society coincides well with our claim; “In order to secure individual claims, people disguise their self-interests by arguing that any failure to reciprocate in a particular instance might establish the precedence of a counternorm of selfishness. Therefore, it is in the interests of the exchange partner to sustain the norms so that he may take advantage himself at a later time.” See also the case of ‘communal food sharing norm’ in Kameda et al. (2005).
greater urbanization and mobility, individualistic lifestyle etc (see Coleman 1982 for this fundamental trend of social change). Nevertheless, the implied principle of reciprocity continues to govern strongly the informal relations of people.

IN WHAT SENSE IS SOCIAL NORM EFFICIENT?: COASE THEOREM RECONSIDERED

Now we want to conclude our discussion by returning to the Coase theorem. We think Coleman’s criticism of the Coase theorem was partly correct and partly wrong. The correct part is that the Coase theorem is wrong and social norms matter. The wrong part is that this is because transaction costs should be assumed as nontrivial or because changed property rights lead to different optimal outcomes. As mentioned before, first, the zero transaction cost assumption, problematic as it is, does not give rise to the genuine problem in the discussion of norm. In the dynamic time perspective taken here, norm could be generated among group members even in the extreme situation of no transaction cost (e.g., within the family) as long as the time consistency and incentive problems remain nontrivial. Admittedly, one of the major functions (or byproducts) of social norm is transaction cost reduction. But relying on its function to explain any social construct is naive functionalism which should be rebutted. We argue that tackling transaction cost is somewhat beside the point here. Second, more seriously, the changed property rights (or the normative constraints) do matter, leading to different social optima by affecting incentives of involved parties. Still, however, the Coase theorem is valid if it is interpreted as a rather vague claim that arbitrage opportunities could be fully saturated through frictionless bargaining toward some social optimum. The theorem does not say that initial distributions of property rights do not matter in achieving a particular social optimum; it only says that initial distributions do not matter in achieving some social optimum, no matter what it may actually be.

How then do social norms (or laws) matter in this Coasian setting? The Coase theorem, it seems to us, does not hold true in certain dynamic settings, and it is in this dynamic perspective that social norm matters in discussing social optimum. If our conception of a norm as a second best optimum can be agreeable, then it must be clear why and where the Coase theorem fails. Social norm, as a set of equilibrium ex ante, constrains people in a systematic way from achieving the first best social optimum, and the specific extent of distortion of social optimality must be agreeable — even if only implicitly so — among group members as a tradeoff for the benefit of mitigating potential incentives to deviate from the prescribed path of conduct in the future. That is, many of mutually beneficial moves among social members, which will certainly contribute to improving social optimality, are simply not carried out because of social norm for the sake of the time consistency of the once agree-on rules.

Let us turn to an illustration from the Confucian norm of “filial piety.” In Confucian
countries, there has been a strong social pressure, especially on the eldest son, to care for aged parents. Let us first consider how such a norm can be socially efficient. Taking the resource allocation problem in a simple version of “overlapping generation” model, we assume that each individual lives only two periods, “young” and “old,” and gives birth to one child in the “young” period; also assume that each can work only in the “young” period. Optimizing his resource allocation, each “young” parent, net of his own and his child’s consumptions, has an option over his residual earnings — either to save as self-support for his old age or to invest in his child whose future productivity, we suppose, increases with such human capital investment. As succinctly shown by Becker (1991), some Pareto improvement over intergeneration is feasible by the parent investing in his child’s human capital and, subsequently, the child repaying a “gift” to the aged parent for his old-age support (when the market interest rate is less than the marginal productivity of human capital — this caveat is assumed to be true here). One obstacle to this Pareto improving move, however, is that the parent would choose the optimal amount of investment in his child only if there were some guarantee for the child’s future repayment to the parent at his old age. The norm of “filial piety” plays the requisite role of social guarantee for the child’s repayment of old-age support to the parent, making it possible to realize the intergenerational Pareto improving move. Then the question is, how is such an implicit contract between parents and children made possible? Suppose the following strategy for each generation; “contribute old-age support and invest in the child if and only if all past generations did the same, or else save for your own and forfeit the investment.” Then such a norm of filial piety would be a Nash equilibrium, and the resulting intergenerational resource movements would become socially efficient. Note that pair-wise mutual contracts from bargaining between parents and children at each point in time would be unlikely to work because the child’s debts cannot be enforced if he deviates once the investment in him have been already made by his parent. Here, contrary to the Coase theorem, the existence of social norm plays a significant role, making a difference in realizing a socially optimal outcome.\(^{17}\) Also note that it is not the transaction cost but the incentive problem that is solved by the social norm. Next, then, how does the second best optimality enter into in this scenario? Imagine the first best outcome, where no incentive problems exist among children (by indoctrination of the child’s preference by the parent, following Becker’s line), and each child-parent pair can freely decide on their own optimal amounts of investment and old age support. Compared with this (incentive-compatible) first best outcome from fully decentralized decisions, the actual outcome from norm following behavior conveys some degree of distortion of social optimality,\(^{17}\) Becker (1991, 1993) focuses rather on the internal process of creating some “guilt feeling” in the child by the parent, the disutility of which motivates the child to pay the old-age support. However, we are a bit skeptical about this approach: If the obligation of repayment is personal in its character, confined to each pair of family members, then why did such a social norm of filial obligation existed as a normative principle and was shared by many over and beyond each individual pair of parent-child?
i.e., the ignorance of household-wise differences of optimal resource flow, because of the normative constraint of social norm for curing the time consistency problem — that is, to socially prevent potential incentives to deviate from the *ex ante* prescribed rule of conduct from arising once actions by one party (i.e., investment of the parent) have been taken. So we return to our conclusion: norms matter in achieving social optimum, and do so only in a constraining sense.

**NORMS IN FLUX**

Even though we have so far treated norms as if they have some distinct phases of evolution, it would be more correct to say that social norms are *continuously* generated, appreciated, observed, and modified at each and every point in time by social members; there are no well-defined distinct phases of evolving norms. Time passes and norms are in flux; whatever once happened has already happened and is not reversible. People have different notions of individual optimality based on their changing social positions at different times, and they have correspondingly different degrees of willingness to accept the prevailing norms at different point in time. The suboptimality of norm following behavior due to time consistency is thus *inherent* in any social norms as far as they are honored in a society (and this is where the Coase theorem fails). Even though they are generated and followed primarily because of their efficiency to coordinate people over some domain of social behavior for mutual benefit, different social norms have different degrees of (distortion of) social optimality depending on the specific context of the society; to have a particular social norm means to choose implicitly a particular degree of social suboptimality by people, who continue, consciously or unconsciously, to exert higher-order discretion over the matter of “rule versus discretion.”

What should be done from here? We do not believe that norm is incompatible with rationality. Nor do we believe that people just blindly follow norms. From our perspective, it is untenable to argue that norm is either a reified entity aloof from individuals (social holism) or just an arational psychological motive residing in the individual mind (psychologism). We contend that a sound understanding of norm must come from a rational reconstruction of it. The task is to “make norm-following as possible come out as rational.”
REFERENCES


Perspectives 14(3): 137-58.
[http://www.econ.jhu.edu/People/Young/PalgraveSocialNormsJuly07JHU.pdf ]

[Submitted October 14, 2008; Accepted November 30, 2008]