

HARRIS SCHOOL WORKING PAPER
SERIES 06.05G

CHAPTER 3: EQUILIBRIUM NORMS *

Howard Margolis

**This is a draft chapter from Cognition and Extended Rational Choice
(forthcoming Routledge 2007).*

Chapter 3 Equilibrium Norms

In one sense a norm is a statistical notion. It is just whatever people on average do. In another sense, a norm is what people are ought to do. We can imagine a community in which people commonly drink and drive, but if we ask about it everyone says it's wrong to drink and drive and they wish the police were stricter so they wouldn't do it. So here, the statistical norm would be entirely different from the "normative" norm. Is there, nevertheless, a connection between the two senses, and if so what does it involve? An inquiry of this sort might reveal nothing more in the case of an interesting word like "love" or "rationality" or "norm" than in the case of an ordinary word, like "foot" or "bank." For either sort of word we will find comparable examples of how the meanings of words evolve by extension and analogy, or merely by migration from another language. However, for norms, if NSNX is sound, it becomes tautological that something deeper is involved. After some definitions and preliminary remarks, I want to consider how NSNX implies a way of thinking about norms, in which there is an essential connection between various senses of the term, and this connection yields insights into how norms change over time, why some behavior comes to be covered by formal regulations backed by police power while other activities are left to voluntary compliance, and so on.

Norms understood as what usually happens characteristically fall short of norms understood as what would people in the society think would be socially best (what a good person would want to do if he could rise above his self-interest). But norms as what usually happens also sometimes go well beyond what a person guided only by self-interest would do. In terms of NSNX, we would look for some "neither selfish nor exploited" balance between these two competing tendencies (what would be best for me, what would be best for all of us). And since norms in either the statistical or normative sense are intrinsically social, like prices in a market, they must emerge from a social process, as prices in a market emerge from a social process. In either case, we get a social equilibrium emerging from the interaction of many individual choices. For prices, this emerges from an interaction among many choices about what to buy and what to sell. For NSNX and norms, the social outcome must grow from individual choices seeking a "neither selfish nor exploited" individual equilibrium.

We want to say something about how that might work.

Figure 1 about here

There are statistical norms for everything -- bowling scores, rainfall, and whatnot, but we are interested in norms of good behavior. We can notice that statistical norms (with respect to good behavior) always look like a compromise between purely self-interested and purely social motivation. And if the NSNX rules govern that compromise, what we see should be the current state of a process in which individuals move toward some balance (equilibrium) in their behavior, which at the social level aggregates to some social equilibrium characteristic of the society as a whole. We want to consider in more detail how that might work.

The categories we need are:

1. *Aspirational norms*, like the Golden Rule;
2. *Pragmatic norms*, which people are expected by their fellows to obey, and which elicit disapproval (or worse) if a person is seen to violate them. This is the primary sense of the word in ordinary usage, and the sense we are principally concerned to explicate here;
3. *Statistical norms*, or what typically is done; and
4. *Personal norms*, by which a person judges his or her own behavior, and which may vary (in any particular case) toward more strictness or toward more laxity from the same person's sense of pragmatic norms.

Norms in the second and third senses (pragmatic and statistical) are irreducibly social. They are characteristics of a society, not of any particular individual. Norms in the personal sense are intrinsically attached to some individual, though always under the influence of social interactions. And a connection between personal and pragmatic norms is inevitable, since the principal factor in what makes one individual see the behavior of another as a violation of social norms must be just that if she were in the same situation she expects she would have behaved better. Someone living in the community would acquire a sense of prevailing *pragmatic norms* which as a first cut might be taken as reflecting the mean, or typical, personal norms in the community.

A clue to the character of pragmatic norms is that, as the reader will be aware, no

one will scowl at you for driving 10 miles over the speed limit when everyone else is doing the same. But the corresponding norm in the statistical sense could be either higher or lower, contingent on the intensity of highway patrols and the scale of fines. So by itself, compliance tells us nothing one way or the other about the *normative* status of a rule. But scowls, sneers and the like tell a lot.

Strictly speaking, *aspirational norms* could also be individual, like personal norms: they represent an individual's judgment about ideal behavior. But they are ordinarily widely shared across a society and, indeed, for the most important of such norms (favoring honesty, fairness, and so on), they are even widely shared across cultures, as indeed must be the case by the criteria for NSNX social preferences Chapter 1. A common remark is that all the major religions have similar moral codes. And although there are certainly major idiosyncrasies across cultures, for many aspirational norms, any society that lacked them would not be a nice place to live, nor a place where cooperative behavior would be sufficiently common to make it a prosperous place to live. On the other hand, plausibly universal norms (like treat equally deserving cases equally, keeping promises, etc.) will be sufficiently general to inevitably run into conflict with other such norms (as when keeping a promise turns out to conflict with treating equally deserving cases equally). Shweder (2003) provides an excellent discussion.

Even an outsider might give a good account of aspirational, pragmatic, and statistical norms familiar to everyone within a community, even though he has never met any individual member (like a scholar who can use documents, myths, and literature to give a good account of norms in some long-vanished society). But even a member of the community, entirely familiar with prevailing aspirational, pragmatic, and statistical norms, might personally regard some of those norms as irrational or perverse. So there can be personal norms which depart a long way from norms in the social sense.

The personal norms of individuals within the community would also be subject to some "local" variation, in the sense that even under fixed external conditions (the environment is not in any way unusual) no one exactly lives up to (or down to) his own standards all the time. In particular an individual who feels at the moment out of NSNX equilibrium will be influenced by that, inclined to do more than usual if the disequilibrium is on the selfish side, and the converse if on the exploited side.

To see how equilibrium norms (in the various senses of "norms" now at hand) might evolve, start by noticing that although we today live in communities that readily sanction self-interested behavior, people nevertheless act as if there were limits to what is *appropriately* self-interested. It is all right to be self-interested but not right to be selfish. On the other hand, there is also a sense that a person can go far enough in the opposite direction to be exploited. It is right to be socially concerned. But it is not wrong to avoid being exploited, even for a good cause.

So there will be the tension NSNX expects between the propensity not to be selfish and the propensity not to be exploited. Ordinarily, social motivation is less conspicuous than self-interested motivation. But the extent of social motivation in ordinary life may be much larger than is commonly recognized, masked in part by the empirical fact that socially motivated choices rarely are absolutely in conflict with private (self-interested) motivation. Even the very best thing socially is rarely an unqualifiedly bad thing privately. And as developed in some detail in SA&R (Ch. 5) NSNX choices will commonly be neither the best thing socially nor the best thing for self-interest but a choice that combines elements of both. Consequently it is often easy to see behavior in terms of self-interest even when very little analysis would be needed to show that the strictly selfish gain is too small to make that plausibly rational. Further, behavior that is social with respect to some community may be perverse and hence be seen as selfish, from the perspective of a wider community, or a different community. Two people may both think: In my group, people respect norms. In your group, people don't know how to behave decently.

But if humans have any social motivation at all (they do not want to be merely selfish), they will have some rational concern about the right way (the socially sensible way) to behave in various circumstances, just as their rational self-interest assures that they will be concerned about the right way (the financially sensible way) to behave in buying a car. On the other hand, if people are also concerned not to be exploited (not to sacrifice much more of private interests for social values than others in similar situations), then they must be concerned about what other people in fact usually do. The first leads to concern about what behavior is most socially valuable (*aspirational* norms); the second leads to *personal* norms and then to an influence on personal norms of how

the rest of the community is behaving (statistical norms), all of which in the aggregate will shape the pragmatic norms of the community.

From the discussion of Chapters 1 and 2, this should have a familiar ring. If we wanted to make all this more explicit, give it foundations, we would be led back to the formal NSNX setup already introduced, as indeed will be done in Chapter 4. Overall, we can expect to see the emergence of personal norms which qualify aspirational norms in the light of statistical norms, but where there obviously will be linkage between the aggregate of personal norms and the pragmatic norms, which in turn will influence statistical norms.

We want a sense of how these relationships might represent an equilibrium outcome (or a tendency toward equilibrium) among the various contending influences at work, most obviously, the selfish versus exploited tension fundamental to NSNX. As in standard economic theory, an individual seeks an equilibrium allocation of her own resources. Here the individual exhibits personal norms that reflect a balance between the competing motives of some degree of social motivation (which favors doing what seems socially best) as against self-interest (which favors ignoring norms except to the extent that either risk of punishment is sufficient to deter violations or promise of reward sufficient to elicit compliance).

For markets social equilibria emerge (prices and quantities) which are characteristics of the overall situation, not of any individual. But the prices that emerge are contingent on individual equilibria governing how budgets are spread across the menu of goods on offer, which in turn are influenced by the social equilibrium (prices). So an individual's equilibrium allocation of spending between, say, bread and beer is contingent on the relative prices, but the prices -- in particular, whether an equilibrium has been reached and which direction things will move if not -- are contingent on the aggregate of individual allocations. For norms, personal norms are individual equilibria; pragmatic and statistical norms are social equilibria (like market prices and supply), with interactions between individual and social equilibria analogous to the connection between individual preferences and social outcomes in the context of markets.

Norms appear in several varieties, some of which may look hard to assimilate to a view which stresses their essential role as promoters of social value. But the most

promising tactic from the NSNX perspective is to take the essential role as fundamental and then see how far the apparently nonfunctional or only weakly functional sorts of norms can be dealt with (on one side) as derivative of the primary function, or (an alternative also consistent with NSNX) as peripheral in some reasonably transparent way to the main argument. In other words, atypical norms might be side-effects which do not conflict with NSNX but are not directly part of a NSNX story. Or they might be actual anomalies for NSNX but sufficiently inconsequential that social motivation within the community would not be mobilized against them. Or they might be actual anomalies for NSNX that indeed challenge the theory. But before considering such complications, I want to work through what I take to be the core case.

From the discussion of criteria and derived rules-of-thumb in Chapter 1 we can specify some aspirational norms that ought to be present in every society, since they are so obviously helpful to promoting social life. Salient examples would be norms favoring telling the truth, keeping promises, and helping those in need, but all focused on members of the community. But from the Darwinian perspective, what counts as my group will be people who emit cues (or are embedded in a context containing cues) that look familiar to me from contexts of social cooperation. In a world of cheap global communications, there is ever more opportunity for "like me" cues to develop, but also more opportunity *for* adversarial cues. Whether "like me" or "not like me" cues are more salient is likely to depend especially on whether in my relation to this person, cooperative or adversarial relations are more salient..

Since it is socially valuable for members of a community to act in accord with aspirational norms, it must be socially valuable to encourage that. In terms of the selfish/exploited tension, people who are too selfish deserve to be punished. Similarly, people conspicuously useful to the community ought to be rewarded. So we should expect "second-order" norms, promoting adherence to primary norms and favoring individuals who seem particularly valuable to the community. But like effort devoted to first-order norms, enforcement and rewarding would be subject to NSNX Rules 1 and 2. So the issues affecting second-order norms are just those involved in assessing primary norms. Questions will turn on how reliably a person is able to recognize behavior that ought to be punished and on the burden accepted in being a punisher. That would

conspicuously include the risk of retribution from the target or even from non-targets who, perhaps mistakenly, perhaps not, see the punishment as excessive or unwarranted, so that pragmatic norms for 2nd-order compliance would be less stringent relative to 1st order norms since they entail this additional element of cost.

But contrary examples can be found when failing to punish involves a *sacrifice* of self-interest because "everyone knows" what is going on, who is being sanctioned, and what is expected to happen (as with shunning). And indeed it is just here that it is easiest to find striking cases of personal norms that diverge from pragmatic norms in a community, where individual at real risk of penalties on themselves may nevertheless help a person the community treats as morally undeserving.

As the NSNX rules make explicit, merely allowing for social motivation does not make self-interest incentives irrelevant. The private incentive comes from the risk that failure to comply will be punished, or that visible compliance will be rewarded. The social component comes from the perceived social value of complying. It is the two together that must be sufficient to overcome the temptation to neglect a norm. Second-order norms say to punish someone seen to unreasonably violate a 1st-order norm. But what defines an "unreasonable" violation? There must be a standard of how good social behavior has to be to be good enough. In the language I am using, there must be pragmatic norms (good-enough-to-pass norms) as well as aspirational norms. In contrast to accounts in which individuals are dichotomized into stable "types" (norm followers and norm neglecters), it would be an embarrassment for NSNX to find many people who either always or never obey norms when evasion would serve their self-interest.

And a final point setting up a NSNX account of what governs the social equilibrium: It would also be an embarrassment for NSNX to find people to be simply incapable of violating a pragmatic norm without some easily detected revelation of their misbehavior (like a person who cannot tell a lie without blushing). Rather, on the account here, such involuntary self-revelation would be tied to violation of the individual's *personal* norms. For if violation of a norm -- perhaps even a norm usually followed by this very individual -- would move a person closer to individual equilibrium than compliance (on this occasion), then it would be implausible (that is, it would make no sense in terms of the Darwinian argument) to suppose that such a person would turn

out to be unable to conceal that noncompliance. Empirically, substantial ability and also some propensity to engage in conscious deception on what seems to the person appropriate occasions seem to be universal. Good poker players are not necessarily persons of bad character though the heart of the game is deception. So on the argument here and also, I think, on the evidence, we would not suppose that compliance turns on something like a simple inability to avoid looking guilty other than when the agent actually *feels* guilty (is ashamed of himself, is plagued by conscience, and so on). Exactly what the notion of personal norms is intended to capture is the standards which a person ordinarily feels guilty not to meet.

However, since pragmatic norms represent qualified, or less demanding, versions of aspirational norms, they intrinsically allow variability in terms of their range of applicability and their relative weight as against other aspirational norms and also in terms of private costs that are appropriately borne in order to comply. Aspirational norms take the form of explicit principles not made relative to the details of circumstance, as in the Ten Commandments. But their associated pragmatic norms depend a good deal on the tacit knowledge of members of a community to recognize how much less demanding a pragmatic norm is relative to its aspirational counterpart. Boundaries are intrinsically fuzzy. So we need to account for how standards so hard to make explicit can actually be effective. Certainly cases occur that leave us unsure, or where we see it one way on one occasion but might see it otherwise on a different occasion, or where someone whose judgment is ordinarily very similar to our own sees the situation one way and we see it another. Still, even though the boundary is fuzzy, we apparently by and large learn to recognize behavior that looks acceptable even though that often falls far short of idealized aspirations. But that is not the only important example of such coordination. The most important example is the ability of members of a community to coordinate on their understanding of *language*, down to subtleties of context-dependent nuance.

A further issue turns on questions of enforceability as discussed a few paragraphs back in the context of second-order norms. Compliance, again, will reflect some mix of self-interest and social motivation, where the self-interest arises because there are often rewards to being perceived as a complier or penalties to the opposite. The most obvious point is that as behavior becomes more anonymous, the effectiveness of given penalties

against noncompliance must fall, and rewards become impractical. Whether you have bothered to vote is usually essentially anonymous. On the other hand, sometimes identifying noncompliers is easy, such as a norm in my summer neighborhood -- understood by all but, so far as I can recall, never explicitly expressed -- that people shouldn't use lawnmowers before ten o'clock on weekend mornings. If someone did not comply, everyone in the neighborhood would notice.

A subtler but essential enforceability issue arises out of ambiguity. Even if behavior is easily observable, are violations well defined? In fact, they are never exactly so, though for some things (the "only after ten" lawnmower norm) the range of ambiguity may be inconsequential. Often, though, that is not so. There are various good reasons for noncompliance: "nobody told me," "I had a prior obligation," "I was sick," and so on, where for the most part each good reason might also be a bad reason -- malingering in one form or another. Further, often norms overlap. It is bad to lie and also bad to harm other people. But sometimes the main effect of telling the truth would be to harm some innocent person. Many less familiar examples could be given. So no one could possibly obey all norms all the time. There will often be some zone of ambiguity about which norm takes precedence when there is conflict. All of this is essential to understanding pragmatic norms.

We ordinarily see a spectrum, not a dichotomy, between compliance and noncompliance. Academics feel an obligation to do some reviewing for journals and presses, but how much is enough? So, aside from anonymity (the chance that no one will notice whether you have complied), there are at least the further possibilities of (1) partial compliance and (2) good, or at least not obviously bad, excuses for noncompliance, which blur the line between socially acceptable and socially perverse behavior. Plainly, all these allow the gradual erosion of a norm. A slide to the left in the compliance curve of figure 1 is easy to understand. But if you look around, for nearly all pragmatic norms, that is not happening. Far more typically, the situation seems stable. Norms are not noticeably either eroding or growing stronger, as the nuanced meaning of language is mostly stable, though also always in some flux. For norms, somehow we also have a mostly stable equilibrium.

To sum up, an empirically plausible view of what norms and compliance are

really like will note, not neglect, that individuals are not dichotomously either norm followers or norm neglecters; that norms themselves -- in particular, pragmatic norms -- are not dichotomously either applicable or nonapplicable to a particular case but sometimes are only more or less, or marginally or debatably, applicable; and that (in addition) compliance itself allows a good deal of ambiguity, as we have just been noticing. Finally, of course, there is the fundamental ambiguity which was the starting point for the whole discussion: a pragmatic norm is intrinsically fuzzy, since it reflects how people in general are expected to behave as judged by individuals "like me" in the society. Although people within a community have a good sense of prevailing pragmatic, there is no great scoreboard in the sky -- no heavenly normative Dow Jones -- that makes available to everyone the *exact* state of norms of the moment.

Indeed, new norms or stricter versions of old norms occasionally appear, as with various environmental norms which have taken hold in recent years. Apparently, people can sometimes be moved to do more as well as be tempted to do less than the prevailing pragmatic norm requires, so that uncoerced compliance does not monotonically slide downhill. In particular, apparently an equilibrium can arise in which the tendency to erosion of a norm is offset by the tendency to comply more than minimally. We want to get a sense of the connection between the individual equilibrium (setting *personal* norms) and the social equilibrium (governing *pragmatic* and *statistical* norms). *Aspirational* norms, being absolute in character, are not part of the equilibrium process, but being socially-shaped they will eventually re-enter the argument.

Consider some just-recognized environmental threat, such that once a person becomes aware of this threat, the intuition is reliably that people ought to work to contain it. But at the moment we are choosing to inquire, people are only beginning to be aware of the situation. Yet some individuals would be more sensitive to this issue (for example, because they believe they know more or because they care more) and some will be in a position to support the norm at relatively low personal cost (for example, because they are richer). Or both. So through an effect on G' , or on S' , or both, the value ratio (G'/S') for these people will be atypically large. From NSNX Rule 1 we would expect such

people to be first to feel some personal responsibility for recognizing the norm even though others are doing nothing. Since others are doing nothing, there is as yet no pragmatic norm. But for some people there is a personal norm to do something about neglect of a socially important opportunity. From Rule 2 we expect compliance to be limited even for individuals who recognize that "something ought to be done" unless opportunities for a favorable impact from their effort are exceptional, or if they happen to be in a position to do something at exceptionally low private cost -- the conditions which would make the value ratio in Rule 1 exceptionally high.

But once some compliance is occurring, others -- not quite such good prospects as the initial recruits -- can more easily be drawn into some degree of compliance, which would (via Rule 2) tend to increase compliance by those already engaged. So there would be positive feedback until some equilibrium level of compliance was reached within the community, balanced between the perceived importance of the issue and the increasing cost of stricter behavior. Eventually people with an unusually high personal norm might start to relax that norm, noticing that they are taking more trouble on this matter than others, though probably not to match the statistical norm (the value ratio would remain higher for them than for most others). We could sketch an analogous story starting from the other extreme, where a once significant pragmatic norm (eg., against premarital sex) has begun to deteriorate. What had been a well-entrenched aspirational norm (it is intrinsically wrong), weakens as a pragmatic norm (sanctions become less serious), and eventually doubt that is even a proper aspirational norm become more widespread. The pill sharply reduced possible costs, which *increased* the net cost of compliance (S' = appeal of sex, foregone by abstention, net of risk of sex).

. Since S' is net cost to self-interest of behavior constrained by following the norm, S' is smaller (other things equal) when risk is higher. So as risk (of unintended pregnancy) decreases, S' (net cost of constraining behavior) increases. Temptation is not as much offset by risk. And this also undermines a source of perceived social value of the norm (its importance for avoiding unintended pregnancies). Hence G' decreases. So G'/S' is weakened top and bottom (the numerator is smaller, the denominator larger) with a now-familiar consequence for sustaining a norm against premarital sex. As G'/S' decreases, the personal norm weakens, which in the aggregate weakens the pragmatic

norm, which always fell well-short of the aspirational norm. And when social consequences of that do not seem terrible, confidence that G' is high for conforming to this norm then begins to erode, as once occurred for women's not smoking and is now far along in the other direction for anyone's smoking. In large parts of society what had been an unquestioned aspirational norm and a much weaker but very significant pragmatic norm turns into a matter of individual choice, like deciding whether or not to take up skiing. . Since S' is net cost to self-interest of behavior constrained by following the norm, S' is smaller (other things equal) when risk is higher. So as risk (of unintended pregnancy) decreases, S' (net cost of constraining behavior) increases. Temptation is not as much offset by risk. And this also undermines a source of perceived social value of the norm (its importance for avoiding unintended pregnancies). Hence G' decreases. So G'/S' is weakened top and bottom (the numerator is smaller, the denominator larger) with a now-familiar consequence for sustaining a norm against premarital sex.

Chapter 4 provides a more formal diagrammatic treatment of the process described here verbally, applicable to norms and many other matters.

But now another puzzle appears, which happens to have the same form as a famous problem for Darwin, though here the solution must be entirely different. Near the outset I mentioned that pragmatic norms might to be taken to be just the typical or average value of personal norms in the community. So in analogy to the blending inheritance problem that beleaguered Darwin, why should not any favorable social mutation be swamped -- have no perceivable effect -- on prevailing pragmatic norms, or at least such a minute effect that any significant shift in norms would take place far too slowly to account for what we see?

Darwin allowed he could not answer what seemed to him the reasonable criticism that the accumulation of slight advantages faced the difficulty that a slight advantage would be diluted to an insignificant advantage as it was transmitted to successive generations. He could only point to evidence that somehow the difficulty was overcome. Mendel's papers were available by then, but like everyone else Darwin did not recognize their significance. But Mendelian (particulate, not averaged) inheritance resolved the

paradox. Here the correction comes from noticing that although the foundation of the underlying NSNX rules is Darwinian, the cultural evolution at issue now is intrinsically Lamarckian, in the rough sense that variation is not random but very often directed (at least some actors are looking for an opening in a certain direction), and can be easily contagious across persons and groups in ways that would make no sense at all in a Darwinian context (since these agents can talk to each other about what would be good and indeed broadcast ideas to others they will never meet). Or using Dawkins' fashionable label, we can notice that memes very commonly propagate and evolve through mechanisms not available to genes.

We have (at least) three possible complementary contributors:

(1) As with the cigarette example, or as with the emergence of new fashions or new ideas of all kinds, this is likely to be a segmented process, with certain sections of a society first affected (and, of course, with the possibility that the novelty never becomes general and perhaps eventually disappears within the lead segment itself). The internal structure (classes, sects, and so on) of a society may not only play a role in the initiation of a change in prevailing norms but occasionally the converse may also be true. Individuals or small groups attracted early to a novelty may now see something more in common with others so moved, so that adherence to certain norms -- even more than with adherence to novel social knowledge of other sorts -- may either reinforce or undermine existing social alliances or create new ones.

(2) Either in the society at large, or within a lead segment, there may be opportunities for magnifying the effect on general perceptions of the choices of a few people. In contemporary societies, this possibility has become especially striking through the influence of television and more recently accentuated by the at will availability of striking imagery through the internet.

(3) Finally, there is the quite different kind of possibility of political action (aiming at a centralized decision of some sort) to push the new norm. This last possibility leads us to the issue of norms supported by rules, meaning here explicit obligations imposed by some authority with penalties for noncompliance. In terms of the NSNX formalism, sanctions (and also rewards) *decrease* the opportunity cost of complying with a norm.

So (underlying and interacting with all these possibilities) a norm of the sort that concerns us here (which requires some subordination of self-interest to group-interest) will usually appear in gradually stronger form, not all at once, starting from shifts in personal norms as described earlier, with the mechanisms just mentioned making it sometimes possible for modest shifts in personal norms to accumulate rather than dissipate.

All these possibilities can be seen more clearly in the diagrammatic analysis coming in Chapter 4.

Suppose that in terms of self-interest, I would prefer ignoring a norm on this occasion even if I were paid \$10 to comply. I might (or might not, contingent on my personal norm and other NSNX-relevant context) ignore the norm. But now I learn that there will be a \$100 fine if my non-compliance is noticed, with a 1% chance that it will be noticed. You still can't say whether I will comply. If I was risk-neutral, the threat would be \$1, but if I am risk-averse the perceived deterrence might be equivalent to quite a bit more. But even without risk aversion, on general economic reasoning (not bound to NSNX) you can only say I am more likely to comply than without the sanction. In terms of the NSNX formalism, S' was at least \$10 without the threat. It now must be less (given the risk, I have to discount my gain from evasion -- or, the same thing, decrease the opportunity cost of not evading), hence whatever G'/S' was before, that ratio now must be larger, making me more likely to comply.

Out of such simple exercises we could elaborate a discussion of principles that might govern -- in fact do seem to govern -- when and how far various norms are left to purely voluntary compliance (supported only by weak, informal sanctions like scowls) and how far norms are made formal rules backed by legal sanctions. If, for example, a norm seemed socially important (G' is high) but also costly to an individual complying (S' is high), compliance would tend to be low, but support for enforcement might be high. Each of us would be tempted to evade, but all of us agree we would be better off if all complied. But effective enforcement would increase compliance, which (by Rule 2, since W would increase) would increase willingness to comply, which would allow enforcement to be focused on the reduced numbers of offenders. We might reach a state

where compliance was nearly universal, and penalties rare but the situation would deteriorate if penalties were rescinded as unnecessary.

Cases obviously could arise -- empirically we know they very often arise -- in which a voluntary equilibrium would leave the society with statistical norms that fall far short of a good social outcome. We have the familiar free-rider dilemma, which is only qualified, not eliminated, by allowing for social as well as private motivation. Where anonymity and ambiguity are endemic, the same factors that keep the voluntary equilibrium too low will make the enforcement costs of securing compliance merely through coercion too high. No large society functions either way (only by voluntary compliance or always through coercion). There are no anarchic societies organized on any substantial scale. On the other hand, horrendous special cases aside, there is always a measure -- usually a very substantial measure -- of what can be called "incomplete coercion" or "partially voluntary compliance", as discussed in Levi 1988 and Margolis 1991.

From Rule 1, people will tend to comply with their own personal norms at some private cost, but the larger the private costs (other things being equal), the weaker those personal norms will be (the narrower the range of conditions will be in which the person behaves the way aspirational norms say a person should aspire to behave). But to the extent that failure to comply can be made costly, then the effective (net) private cost of compliance decreases. As has been noticed several times, the private gain from a violation is not the gain from evasion without qualification but only the gain after discounting for the risks of evasion. So, of course, the amount of social motivation required to secure any given level of compliance falls as the effectiveness of coercion increases.

What is not so obvious, but has now been inserted into the discussion, is that pragmatic norms, not merely statistical norms, may be strengthened by coercion. Under appropriate conditions norms and rules can interact such that the norm becomes stronger, which makes the rule easier to enforce. The most important conditions are that both the nature of the coercion and the behavior mandated by the rule are seen as supporting aspirational norms. If so, the costs of policing (the costs of effectively imposing the rule) may decline over time -- to some equilibrium level, not to zero -- because pragmatic

norms are becoming stronger.

As an example of the several ways in which that favorable interaction might occur, consider the case of socially useful behavior resisted simply out of the common tendency to feel uneasy with behavior that runs contrary to what we are accustomed to seeing. In a moment I will say something about this tendency to conform. The puzzle there is to see how mere conformity would fit into the general argument about compliance with norms. But that the propensity to conform exists is empirically obvious. Given that it exists, then simply making some novel behavior commonplace and familiar may be very important, even if at first compliance is mandated by a rule. The U.S. civil rights legislation of the 1960s appears to be a case where enforced breaking of customary practices sharply accelerated an important shift in social norms.

The possibility of feedback between rules and norms (coercion and voluntary compliance), of which I have mentioned only the simplest mechanism, illustrates the complexity of a thorough analysis of equilibrium norms. We are trying to understand societies with complicated social structure, subgroups with subnorms, conflicting loyalties, and, of course, criminals and at least a scattering of psychotics. Yet the core process is not itself complicated, though even a reasonably simple core account will seem complicated until some familiarity with it is in hand.

Individual equilibrium with respect to personal norms, governed by NSNX Rules 1 and 2, should ordinarily lead to the appearance of (intrinsically social) pragmatic norms. But pragmatic norms tautologically cannot be so strong that they fail to secure compliance across a group that a person holding a personal version of that norm sees as "people like me". Perceptions of social value will play a large role in the story (from Rule 1), as will perceptions of what others in the community are doing (from Rule 2). Overall, equilibrium norms will be socially important even when not backed by rules (enforcement). Norms weak enough to secure general compliance even without legal sanctions do not always mean not strong enough to be important. But (especially) when a socially adequate norm would involve large private costs, we expect to see norms bundled with rules (voluntary compliance bundled with coerced compliance).

I have been considering here only partial equilibrium at the individual level, taking commitment to the norm under discussion as independent of other social

commitments by the individual. That is all right for this preliminary sketch. But there must be many circumstances where neglecting those interactions would be misleading, since the aggregate burden on self-interest of complying with pragmatic norms that hold within the various groups that a person in some contexts sees as "people like me" must eventually have an effect on inclinations. The aggregate is not fixed. If everyone is doing more, everyone will feel more comfortable about doing more. But the expansion of commitment is not unlimited.

On the logic of the Darwinian argument, norms apply to dealings that affect others the person sees as part of his community or, with qualifications, to others who might have cooperative interactions with the community. So as suggested earlier, what is cognitively seen as the community, therefore, is an important aspect of understanding norms, and a difficult one. It is consistent with the Darwinian argument and easily observed that for individuals living in large, complex societies, what is seen as the relevant community is context dependent. The relevant community is sometimes seen as wider than would ordinarily be expected (extending far beyond the immediate community). But boundaries may also turn out to lie between classes or sects within what an outsider might consider one community. Two people who see themselves as part of a shared community in some context may see themselves as part of adversarial communities in another context, with consequences for what norms seem binding on their interactions.

As with scientific and historical or any other knowledge that reflects what people have learned on a wide scale (across the experience of many people), knowledge of norms will ordinarily have a dominant social component. For all but narrowly personal knowledge, the primary determinant of what a person feels he knows is simply what it seems that "everybody knows" in his society. In their most fundamental significance, aspirational norms are rules of thumb about what "everyone knows" would be socially best. Pragmatic norms are what "everyone knows" is socially acceptable.

Articulation of the social character of knowledge (of how it comes about that intuitions, habits of mind, and so on are coordinated so that "everyone knows" certain things) is not something to be pursued peculiarly in the context of norms. The paradigmatic context already mentioned is the evolution of language. To a remarkable

degree the members of a language community fluently understand each other down to fine shades of meaning, though for words as well as norms, there is no billboard in the sky to account for that remarkable sharing of meaning. We have dictionaries, but it is unlikely that you will fully catch exactly the intended sense of a word by looking it up in a dictionary. And of course people with no access to dictionaries can communicate with subtle nuances of meaning anyway.

Empirically, no one can doubt that coordination of intuitions does occur. Somehow common notions, meanings, gestures, and so on become very well established. They are what "everyone knows" who is a member of the community and, indeed, not to know them, unless you are an outsider, is taken to be a sign of some mental affliction. The process of becoming established as something that "everyone knows" is not beyond analysis. But for the present purpose, it is sufficient to notice that such a process operates for many categories of social knowledge and there is no reason to suppose that something unique is required to account for the social character of knowledge of pragmatic norms. Rather, noticing that norms provide only one of many examples of the coordination of intuitions across a community lets us see that a secondary role of norms could arise as the coordinator of merely conventional social rules regarding dress, manners, gift giving, and so on.

The propensity for coordination that facilitates socially critical functions, such as communication and the division of labor, helps account for a taste for coordination even on matters, such as dress, where no essential functional point seems to be at stake. But even on such mere matters of convention, we should not neglect the extent to which such matters contribute to mutual identification of conformers as members of a community or at least have roots of that sort. The point is important, since we have already noticed that what an individual sees as the relevant community is often crucial to the judgment of what to do.

On the NSNX argument, the tendency to prefer anticipatable, familiar behavior makes some sense. For it is important for coordinating expectations, communication, and so on with respect to language and many other sorts of intuitions other than those about norms. Norms would not have a sharp boundary but, rather, would tend to carry over to other matters. Hence, we can get pragmatic norms without an immediate link to

aspirational norms, where what makes the norm socially valued is not some functional contribution to social efficiency but only the looser propensity to value social coordination for its own sake. We get a norm that amounts to: "unless you have good reason not to, do the usual thing, since it easily causes trouble when unexpected things start happening."

So we get norms that say you ought to dress in the usual way for various social occasions, with "the usual way" varying from context to context. The essential distinction to be made here is between customs (the usual thing) and norms (the right thing), which mostly but not always coincide. Tipping customs vary from context to context; the norm is to leave the customary tip. Similarly, if coordination is seen as usually beneficial, then the norm is to do the customary thing unless there is good reason not to, where the final proviso makes this a pragmatic rather than aspirational norm.

An instructive example on the possibility of selling a place in line. That is just not done. There is no explicit rule against it. Few people have even thought about it. But there was no explicit rule about lawn-mowing before 10 in my summer neighborhood. It just is not done, and if you ask about it people who have never given the matter a thought readily know that "everyone knows" it isn't done. What of people with a high value of time? Wouldn't it be efficient to let them trade money for time with people who would like to do that? So does this not violate NSNX expectation that norms will favor social efficiency? And in a case like this the equilibrium process I have been describing doesn't explain where the norm comes from. But for NSNX or any other plausible view, norms favor efficiency but also other social valuable concerns such as fairness and morale. So there are contexts where buying and selling is what we want to be able to do (conspicuously, a market), but ordinarily we do not want to be bothered with that. I mow my own lawn, but I would not welcome a neighbor inquiring how much I would want to mow his lawn, or take out his garbage, or rent my wife.

It is socially useful to have conventions (common understandings) of where naked cash is all that counts, and where offers to buy and sell are appropriate. And treating a place in line as a marketable commodity would create a socially wasteful occupation of standing in line with no intention of buying a ticket, but only intending to extract some money from someone who would be further back in the line. And why not when waiting

in any other line, or sitting in a bus when there are standees, and endless further cases?

So it is not surprising that custom develops about when buying and selling is socially appropriate, supported by a general norm against upsetting social conventions without sufficient reason. Without spelling out more, it is easy enough to see that although a norm against a place in line being treated as for sale does not immediately look like the result of the equilibrium process I've sketched, it is easy enough to see just such an equilibrium process yielding social understanding of where the line might be that sets off contexts where offers to buy and sell are appropriate and where they aren't, with selling places in line clearly on the "not done" side of that line.

Another issue is that for the equilibrium process to work, an individual must be able to get some sense of the social value of compliance. The emergence of social knowledge must start with individual judgments. So how might a person get a sense of the social value of his individual behavior when he is only one of millions of people in a large society? But this is probably not so very different from the same issue for a member of a small hunter-gatherer band. For many social acts, even within a small community, involve effects too small to be individually perceptible, as, indeed, is also the case for many purely private acts. What is the effect of taking a vitamin pill today?

Such questions often yield confident intuitions if asked on a larger scale: is it worthwhile taking the vitamin pill as a habitual matter? If I am in a position to make that judgment, it is not surprising that my intuition is that it is worthwhile today, even though on any particular day (like today) the bother may be noticeable while the benefit from that day's effort is wholly imperceptible. Similarly, with social judgments it will often be easy enough to reach confident intuitions if the Kant-like question of is "what if everyone in my position did this?" For both the private and social judgments, "seeing" the question in terms of a larger-scale question, where personal experience and common knowledge within the society can be brought to bear, makes things far more manageable. Russell Hardin has pointed out to me that the problem is complicated by the possibility that what would be good if we all are doing it may be worthless if I alone am acting. The point will play an important role in Chapter 10. And the converse case can also arise, where there would be no particular value in my acting if the rest of you are already doing your share,

but if you are not, there may be some special social value in someone's taking the initiative. So the point of this paragraph is not the end of a discussion, but it is a large start, and as much as needed for the discussion here to return to the question of what defines a pragmatic norm. It is that violations arouse resentment in others. But how does a person judge whether another person's behavior warrants resentment?

In thinking of how to answer that question about pragmatic norms, which is essentially a question about what is ordinarily regarded as excessively selfish behavior in the community, we can start by asking how a person recognizes when his own behavior would be too selfish. For the two questions (the first about the self, the second about others) may amount to much the same question. A person has a sense of what feels neither selfish nor exploited with respect to himself (personal norms), which yields intuitions about how another person ought to behave in terms of "how would I behave if I were in that situation." This amounts to supposing that moral resentment remains quiet provided others behave about as well as we expect we would behave in that context. But we are offended to see others fail to do that. Pragmatic norms then emerge from the experience of individuals about what arouses resentment (provokes scowls, nasty remarks, and, occasionally, worse) in the many interactions a person encounters (not just those involving himself) in which people look with approval or disapproval on others' behavior. Empirically, that is a powerful influence on choice, commented on (for examples) by Maynard Smith and by Robert Sugden (1988), by Jon Elster (1989) and in Chapter 8 of SA&R.

But there will also be "second-order scowls," where A is offended but B responds with a look that reflects a sense that A's scowl is unwarranted, and it is out of the aggregate of such interactions that a generally shared sense of the pragmatic norm emerges. So there is (of course) no place where you can look up what the pragmatic norm is in this situation. Rather, there is a social process out of which emerges a roughly shared sense of what looks like appropriate behavior, akin to the many other cases of shared social intuitions discussed earlier. In a completely homogeneous society, equilibrium *pragmatic* norms would straightforwardly determine how far *statistical* norms go beyond what is coerced. Even in large and complex societies, within a reasonably homogeneous subgroup we should expect almost the same. In general, we can

expect a strong propensity (shared with many other aspects of social knowledge) to follow the usual way of seeing things, though, as in the earlier discussion, individual variability is always present and critical in accounting for the possibility of changes in prevailing norms..

On this account, the effect that a mere glance of disapproval can have (even a waiter's imagined scowl on finding a stingy tip after you have left) should not be understood as a punishment effectively inflicted. If a person believed that a certain pragmatic norm was, in fact, socially perverse (as, for example, a white liberal in the old South or, even more obviously, a black activist there), scowls of disapproval would hardly be painful. But when a person's personal norms with respect to whatever is at issue are not radical or rebellious, then the person ordinarily wants to do the usual thing: it is ordinarily what leaves him comfortable (in individual equilibrium). Consequently, a person would be looking for indications of what people around him see as reasonable or unreasonable behavior. Hence, the latent function of the scowls is to signal a target (and others) that the behavior does not appear good enough (or, in the opposite case, smiles function to signal that the behavior is to be admired). After all, what could any effect achieved by the scowl consist of unless the object of the scowl really wanted to behave in a way that would ordinarily avoid scowls?

But, as with many other issues touched on in this survey, the point is not that the deep question can be settled with a quick remark but only to suggest how the NSNX point of view I have been sketching might be able to yield principled answers to a wide range of such puzzles.