Norms and Games

Russell Hardin†

There are three centrally important ways in which norms have been elaborated and explained: in terms of religious or natural law strictures on behavior, in terms of constraints imposed by rationality, and, recently, in terms of agents' behavior in well-defined games. The principal difficulty of a gaming account of norms is to show how the account explains motivations of individuals to follow the norms. This issue is examined in the context of small-number norms and large-number norms.

There are three centrally important ways in which norms have been elaborated and explained. In the oldest literature they are supposed to arise from more or less religious or natural law strictures on behavior, although the terminology of norms is not part of this literature. One might refer here to moral rules, which ought to be followed. In fairness, however, the traditional literature on norms is vast and varied, so much so as to seem ad hoc. For example, norms are even explained as the product of psychoanalytic processes (Scott 1971), as if they were the natural creation of the superego. In the more recent literature of rational choice and the explanation of behaviors from the interests of actors, norms are often functionally explained by the interests that they serve.

The rational-choice theory of norms is a relatively new direction, with early contributions from David Lewis (1969) and Edna Ullmann-Margalit (1977), although one can find elements of such a theory in many earlier works (e.g., Weber [1904–5] 1958 and many places in Hume [1739–40] 2000, Book 3). In this second body of work, norms are not necessarily seen as good or bad; they are neutral. Indeed, a norm that might serve my interests might harm yours. Or a norm that once served a society well might now be residue that is harmful.

The third literature is still more recent and is an outgrowth of experimental games. Cristina Bicchieri, in her strongly argued book *The Grammar of Society* (2006), adds a major new chapter to this branch of the

†To contact the author, please write to: Department of Politics, New York University, 19 W. 4th Street, New York, NY 10012; e-mail: russell.hardin@nyu.edu.

Philosophy of Science, 75 (December 2008) pp. 843–849. 0031-8248/2008/7505-0029$10.00
Copyright 2008 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/594528 Published online by Cambridge University Press

theory of norms. This body of work typically tracks behavior in some well-defined game, manipulates background conditions of the play of the games, and reaches conclusions by beginning with behavior and inferred motivations. The rational-choice accounts begin with theory, and the traditional literature usually begins with a specifically normative, prescriptive theory. Neither the rational-choice nor what I will call the gaming literatures is prescriptive.

In the rational-choice literature, what motivates individuals to follow norms is supposed to be their interests in achieving the results of norm-following or avoiding sanctions. In the quasi-religious literature, what motivates is the hand of god or nature, who will slap us if we behave wrongly and fail to follow the norms. In the recent gaming literature, what motivates is in question. Some experiments find that identity plays a very large role while others focus on reciprocity as a supranorm that guides lower level norms that guide action. Psychological anomalies also play roles here. Many of the game experiments on norms attempt to separate the effects of a desire for reciprocity and those of identity in motivating players to follow certain norms.

In general, the rational-choice and the gaming accounts of norms focus on the explanation of norm-following, while the quasi-religious and moral accounts do not heavily worry about the explanation of behavior but more about judgment and evaluation of it. I will therefore compare the gaming approach only to the other systematic explanatory approach. One might say that the chief difference between Bicchieri's gaming approach and the rational-choice approach is that the former gives far more play to a diversity of psychological accounts of various motivations. This fact already suggests a characteristic difference between these two approaches. The rational-choice approach is highly systematic and focused on a single motivating consideration—interests—without concern with the niceties of more complex and complicated psychological phenomena. Those in the gaming approach might harshly criticize the seeming psychological simplification of rational-choice theory. They might even suggest that the rational-choice crowd is sadly unable even to keep up with the manifold intricacies of a rich psychological account of norms. And they might well be right in this final assessment—they are right about me, for example—but the rational-choice theorists would generally think that this is not an important failing.

To make the rational-choice account of norms work, there must be a mechanism that punishes or otherwise sanctions violations of the norms. In general, rational-choice theorists would want to find a form of endogenous sanctioning in which the sanctioning itself directly serves the interests of the sanctioners. This is easy for norms governing certain small-number interactions such as promise keeping. For large-number inter-

https://doi.org/10.1086/594528 Published online by Cambridge University Press

actions, rational-choice theorists would want to show that there is a perhaps invisible and subtle reaction to violations, independent of the intentions of the rest of the community, that benefits from maintenance of the norm. As I will spell out below, this latter vision fits the functional explanation as developed by Robert Merton (1968) and Jon Elster (1979), among others (see also Hardin 1995, Chapter 4). This functional explanation model fits certain large-number norms, as I will show below.

1. Norms in Small-Number Relationships. Small-number norms are relatively easily explained from rational-choice principles. Consider one small-number norm: promise keeping. Bicchieri discusses this norm more or less only in passing and assumes its force in many instances of supporting or complying with other norms. One might find this surprising since the norm of promise keeping is possibly the most commonly cited norm in all of the literature, especially the literature of twentieth-century philosophy. But neglecting this norm is not surprising in an important way because it is hard to make sense of such a norm in one-off relationships, which is the modal structure of the gaming literature. One might suppose that any theory of norms that cannot handle promise keeping is a nonstarter, or one could insist that promise keeping is somehow embedded in many of the cooperation games that generate claims about norms. I think the latter claim is likely true, but if so, then we should want to see a direct effort by the gamers to explain promise keeping.

There is a class of important universalistic norms that is virtually self-enforcing in many contexts. These are norms such as those for telling the truth, keeping promises, and maintaining fidelity to spouses and friends. They are self-enforcing when they govern ongoing relationships between pairs or very small numbers of people. In brief, I will refer to these as *dyadic norms*. The enforcement of these norms comes naturally from the fact that *the relationships that they govern are of value to the participants* beyond the instant interaction on which someone must keep the relevant norm or violate it. These characteristics of the relationship suggest that it is an iterated prisoner's dilemma in its incentive structure. Each participant in the dyad sometimes has a short-term interest in violating the norm but a long-term interest in maintaining the relationship. If the latter is great enough, it can trump the short-term interest and make it worthwhile to forgo short-term gains in the interest of longer term gains (see Hardin 1988, Chapter 2).

One might tell the truth out of strict moral scruples, but one also has an interest in telling the truth to one's ongoing relations. Even if you are bound by moral scruples, you may nevertheless depend on the incentives of self-interest that keep others honest enough to make dealing with them worth your while. The usual resolution of an iterated prisoner's dilemma

https://doi.org/10.1086/594528 Published online by Cambridge University Press

in which one party fails to cooperate often enough is not to join that party in cheating but simply to withdraw from the relationship. That can be a brutal sanction. If others have an interest in cooperation with you, you have better prospects from cooperating than from withdrawing or cheating. In general, you have better life prospects if those around you recognize their interest in maintaining cooperative relations with you (Hardin 2002, Chapter 1). Being cooperative loses much of its value if too few others are cooperative, and it may finally even become disvalued. Similarly, truth telling may often seem to serve perverse purposes for many relationships in a Nazi or other totalitarian society. Typically, then, these dyadic norms are straightforwardly self-enforcing and not inherently moral. Such a norm may govern a very large population, all of whom are involved with various others in dyadic relationships. But there is no large-number equivalent of the dyadic norms. Many, perhaps most, the game experiments focus on dyads.

2. Norms in Large-Number Relationships. A norm that governs relationships that are inherently large-number rather than dyadic cannot be reinforced merely by the withdrawal of particular cooperative parties from interactions with uncooperative parties (Hardin 1982, 153–154). Norms of difference and exclusion govern large-number relationships of identification of a whole community and exclusion of others from it. They are reinforced through the mechanism implicit in their functional structure, not directly through the iterated incentive offered by the fellow group members in a large-number prisoner's dilemma. The explanation of large-number norms in general cannot turn on a rational regulation of the problem of collective action in enforcing the norm, although there may be some cases that can be regulated through somewhat unstable conventions (Hardin 1982, 173–205).

Norms that govern large-number interactions are therefore somewhat more difficult than dyadic norms for, say, promise keeping. If one of these governs a large-number prisoner's dilemma, then it would seem to be subject to free riding that would ruin the effectiveness of the norm. What we need is a system for motivating cooperative behavior in such an interaction. Many larger-number norms can readily be fitted to rational-choice principles through a simple feedback model of functional explanation. For example, suppose we wish to assert that a norm of loyalty to a group exists. We might crudely suppose that it exists somehow because it supports, say, intragroup cooperation. That is to say, the norm is functional for cooperation. Now we can explain the maintenance of the norm (but maybe not its origin) by the paradigm of functional explanation.

An institution or a behavioral pattern X is explained by its function F for group G if and only if:

https://doi.org/10.1086/594528 Published online by Cambridge University Press

1. *F* is an *effect* of *X*;
2. *F* is *beneficial* for *G*;
3. *F* maintains *X* by a causal *feedback* loop passing through *G*.

The pattern *X* here is following the norm of basic loyalty to a group; its function *F* is social cooperation (or other related concerns) for the group.

1. Now cooperation (*F*) is an effect of following the norm (*X*), that is, cooperation is built up of actions of keeping to the norm.
2. Cooperation (*F*) is good for—serves the interests of—the members of the group (*G*).
3. Their cooperation (*F*) maintains their loyalty (*X*) by a feedback loop passing through the members of the society (*G*).

Let us elaborate on step 2 in this explanation. Their cooperation is good for the members of the group. It would be wrong to say that the norm of group loyalty is a norm that serves our interests generally or that it is inherently good. This functional account works for many groups, including morally awful groups such as the Ku Klux Klan or a street gang. It also works for church groups, academic seminars, and so forth.

A large fraction of all standard large-number norms take this form. But that means that they are based in communal relationships. They cannot be readily studied in contexts that abstract from behavior in an ongoing set of relationships. Laboratory games often stipulate conditions that set aside all of this standard context that actually defines and enforces standard norms on the functional, rational-choice account. Gamers might suppose that players invent norms in context. Bicchieri generally supposes that the issues and contexts of her games evoke already extant norms, such as for fairness, to develop a more specific norm. Does this, as Bicchieri says, turn a large-number prisoner's dilemma into a coordination interaction? If I could cheat—be a snitch, for example—it is possibly a prisoner's dilemma. What would make it a coordination game?

3. Motivation. Bicchieri's condition b for a social norm to exist in situations of type S is: "Normative expectations: i believes that a sufficiently large subset of P (the relevant population) expects i to conform to [norm] R in situations of type S." This seems to make i actually "prefer" to conform to the norm (2006, 11). This preference is not well argued. If we invoke Bicchieri's second version of this condition b′—the version that includes sanctioning for noncompliance—we are left with the need to explain the endogenous and presumably spontaneous sanctioning by others in the group or society. Condition b′ is stronger: "Normative expectations: i believes that a sufficiently large subset of P (the relevant pop-

https://doi.org/10.1086/594528 Published online by Cambridge University Press

ulation) expects i to conform to R in situations of type S, prefers i to conform, and may sanction behavior." In this version of the condition, we may have motivation to conform from sanctions, although there is no account of why anyone would sanction someone not complying with a norm.

Note an implication of condition b, which does not include sanctions. Here i must be normatively committed to following the norm. The claim seems to be that if there is a norm, people are normatively committed to following it. Either this is a trick of our psychology, or motivation, which one might have thought we must explain, has been folded into the definition of a norm. Others' expectations of what I should do will motivate me to do what they expect of me. I think this claim is true for a class of expectations that Bicchieri does not count as norms. Simple coordinations, such as the driving convention of driving to the right (or the alternative convention of driving to the left) are not instances of Bicchieri's "situations of type S, because S can be represented as a mixed-motive game" (2006, 11).[1] In coordination interactions, we have strong internal motivation (of acting in our own interest) to follow the convention in place where we are. Bicchieri calls such conventions "descriptive norms."

It is reasonable for Bicchieri to differentiate descriptive from her proper norms because the latter help to overcome one's interest in doing other than what the norm calls for. Without the sanctions of condition b′, there is no such easy internal motivation as there is for following descriptive or coordination norms. I always urgently want to follow the driving convention and I have no competing desire to cheat on that convention. But in many contexts I might have a competing desire to violate the norm of truth telling or promise keeping. Hence, we need an account of why the desire to follow the relevant norm overcomes any competing desire to violate it. Bicchieri says, if condition b is met, then "*as a result*, i prefers to conform to R" (emphasis added). Why? Without a rigorous answer to this question, the treatment of norms here is partially a throwback to the oldest—religious and moral—literature on norms, in which normative commitments are simply stipulated.

4. Concluding Remarks. The principal difficulty of a gaming account of norms is to show how the account explains motivations of individuals to follow the norms. Bicchieri often explains compliance with norms as the

1. Ullmann-Margalit (1977) counts conventions that resolve coordination problems as norms. A coordination game presents a problem when there are two or more equally good coordination points or strategies. On her account, a national driving convention is a norm—or at least once was. The difference here is perhaps just definitional, although Bicchieri makes it do explanatory work in her model.

https://doi.org/10.1086/594528 Published online by Cambridge University Press

result of psychological anomalies or mistaken perceptions about the world and others' likely intentions (2006, 182ff.). It is the first concern of a rational-choice account to motivate individuals, but the rational-choice model of motivation is relatively simplistic. An advantage of this simplicity is that we do not need ad hoc stories of why a particular psychological anomaly is at work in some behavior. But maybe the potentiality for ad hoc storytelling is an implicit risk for an adequate account.

REFERENCES

Bicchieri, Cristina (2006), *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.
Elster, Jon (1979), *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge: Cambridge University Press.
Hardin, Russell (1982), *Collective Action*. Baltimore: Johns Hopkins University Press/Resources for the Future.
——— (1988), *Morality within the Limits of Reason*. Chicago: University of Chicago Press.
——— (1995), *One for All: The Logic of Group Conflict*. Princeton, NJ: Princeton University Press.
——— (2002), *Trust and Trustworthiness*. New York: Russell Sage.
Hume, David. ([1739–40] 2000), *A Treatise of Human Nature*. Reprint. Edited by David Fate Norton and Mary J. Norton. Oxford: Oxford University Press.
Lewis, David K. (1969), *Convention*. Cambridge, MA: Harvard University Press.
Merton, Robert K. (1968), *Social Theory and Social Structure*. Enlarged edition. New York: Macmillan.
Scott, John Finley (1971), *Internalization of Norms: A Sociological Theory of Moral Commitment*. Englewood Cliffs, NJ: Prentice-Hall.
Ullmann-Margalit, Edna (1977), *The Emergence of Norms*. Oxford: Oxford University Press.
Weber, Max ([1904–5] 1958), *The Protestant Ethic and the Spirit of Capitalism*. Reprint. Translated by Talcott Parsons. New York: Scribner's.

https://doi.org/10.1086/594528 Published online by Cambridge University Press