Where Do Norms Come From?

By CASS R. SUNSTEIN

Review of THE EMERGENCE OF NORMS, by Edna Ullmann-Margalit

Oxford University Press, 1978

Before her premature death in 2010, Edna Ullmann-Margalit produced a series of elegant and highly original essays, focused about all on rationality and its limits. Hers is the best philosophical work on invisible-hand explanations and presumptions. With Sidney Morgenbesser, she is responsible for the important distinction between choosing (deciding on the basis of reasons) and picking (as in flipping a coin). In her later years, she examined how rational-choice theory might be defeated when people make Big Decisions, where their values and even their character are on the line. Ullmann-Margalit also explored, with grace and sensitivity, the idea of considerateness and its role both in daily life and within the family. (This may be her finest work.) None of her essays shouts from the rooftops, but all of them leave large subjects richer than they were before -- and some of them create new subjects altogether. They have a timeless quality, and they also have a kind of synergy; at some point, they ought to be collected into a single volume.

In the meantime, Oxford University Press has just published the first paperback edition of Ullmann-Margalit's The Emergence of Norms (1977), which has a strong claim to having spurred the last decades' outpouring of work on that topic. Almost four decades later, the book repays careful reading, not least (I think) because of the discussion of norms of partiality, which has been much neglected, and which raises on a number of unresolved problems of social theory.

Ullmann-Margalit sees her work as "an essay in speculative sociology" (p. 1), designed to understand both the rise and the function of social norms. She describes her thesis as this: "certain types of norms are possible solutions to problems posed by certain types of social interaction situations" (p. 1). Her claim is one of "rational reconstruction," involving not historical evidence, or indeed anything at all empirical, but instead a plausible claim of how a practice or a phenomenon might have emerged. She treats three such situations as paradigmatic or "core," involving prisoner's dilemmas, coordination, and inequality. In all three situations, familiar social norms turn out to resolve the specified problems. In her view, norms typically have that effect insofar as they impose "a significant social pressure for conformity and against deviation," alongside a "belief by the people concerned in their indispensability for the proper functioning of society," and an expectation of clashes between the dictates of norms "on the one hand and personal interests and desires on the other" (p. 13).

With respect to prisoner's dilemma situations, the central argument is straightforward. If people are facing such situations, they cannot easily produce a mutually beneficial state of affairs without "a norm, backed by appropriate sanctions" (p. 22). Suppose, for example, that the question is whether to pay one's income tax, to vote in a general election, to keep a promise, or to cut through a neighbor's well-tended lawn.

In each case, a PD norm might turn out to be especially important, all the more so "the larger and the more indeterminate the class of participants, and the more frequent the occurrence of the dilemma among them" (p. 25).

In short, norms operate as stabilizing devices. Ullmann-Margalit understands the logically faulty but widespread generalization argument ("what if everyone did that?") as a reflection of the psychological power and hence the utility of PD norms. She also emphasizes the important but underappreciated fact that it sometimes make sense for norms or law to keep people in PD situation; consider the antitrust laws (pp. 44-45).

Though Ullmann-Margalit devotes a lot of pages to coordination norms, they are a sense much simpler, because the interests of the parties coincide. In the pure version, we both (or all) want to meet somewhere in New York City; the question is where. A coordination norm tells us to meet at Grand Central Station. In the non-pure version, our convergence of interests is not exactly perfect (maybe Penn Station is a bit easier for you), but it is close enough, because we care more about coordinating than we do about getting our way on exactly where.

Ullmann-Margalit urges that in recurring coordination problems, people tend to arrive at a successful solution, which then becomes a norm. Familiar examples include norms involving dress, the acceptance of legal tender (and hence, perhaps, the rise of money), driving on the left (or right), and etiquette. In novel coordination problems, bottom-up development of norms is not possible, and hence "a solution is likely to be dictated by a norm issued specifically for that purpose by some authority" (p. 83). In either case, she insists that a coordination norm is no mere regularity of behavior. It is "supported by social pressure" and in that sense "it even slightly changes the corresponding pay-off matrix" by making a "particular coordination equilibrium a somewhat more worthwhile outcome to be aimed at than it would otherwise have been" (p. 87). Here she understates; if a norm is supported by a lot of social pressure, it might do far more than "slightly" change people's payoffs.

In an instructive brief discussion, Ullmann-Margalit urges that for the individual conformist, all decisions are, in essence, coordination problems, in the sense that the goal is "to meet the others in their choices" (p. 93). From that point of view, the conformist faces "a unilateral coordination problem" (p. 94). We can even imagine "a society composed entirely of conformists," whose members think that conformity matters more than anything else, thus raising the question whether, in such a society, any decision at all might turn out to be a coordination problem. The thought experiment is interesting, because some groups are a lot like that. Ullmann-Margalit urges that if "it is common knowledge in this society that they are all avowed conformists who are, moreover, content to act that way," then coordination problems are indeed pervasive.

From the standpoint of the contemporary reader, her most novel and challenging (and alas far from entirely satisfying) discussion involves norms of partiality. Ullmann-Margalit begins by assuming a status quo of inequality, "such that one party is more favourably placed than the other" (p. 134). As she sets up the situation,, the disfavorably placed party is quite aware of being disfavorably placed and wants to improve his position. She emphasizes that the goal might be to improve either absolute position (for

2

example, by having more opportunities or resources) or relative position (by narrowing the gap with the other party, holding absolute position constant). By contrast, the favorably placed party wishes to maintain the status quo.

In Ullmann-Margalit's account, the most interesting problems, for these particular players, have two features. First, the status quo is in game-theoretical equilibrium, in the sense that neither side can improve his absolute position by a unilateral deviance. Think, for example, of a relationship between employers and employees, or husbands and wives. Second, the status quo is strategically unstable, in the sense that the disfavorably placed party might be able to improve his relative position, but sacrifice his absolute solution (at least in the short-run), by a unilateral move (or rebellion). Think, for example, of a threat to strike or to leave the marriage. We can imagine many real-world analogues, including not only labor rights and the division of labor within the family, but also the wide variety of groups that consider whether to engage in civil disobedience.

Ullmann-Margalit's question is this: How can the favored party try to stabilize the status quo? There are a lot of possibilities. One, of course, is to use force. Another is "to share some of the benefits of his favored positions with the other party" (p. 169). Another is to conceal and blur his favored position (an especially interesting and potentially successful strategy). Another is to try to exclude himself from the frame of reference used by the disfavored part. Yet another is to take steps to convince the disfavored party that he is, in fact, much better off than he was in the past (or might be under some other arrangement). All of these strategies are familiar from past movements for equality, and all have both potential and risks.

But Ullmann-Margalit's particular interest is in the use of norms of partiality, which operate to stabilize otherwise volatile situations. As examples, she points in particular to norms associated with property, including prohibitions on trespass and "the inheritance institution," which she says are meant "to preserve, protect, and perpetuate the position of the 'haves' – and their descendants – in states which are inherently states of inequality" (p. 173). She acknowledges that norms of this kind might take the form of law or instead customs. But she insists that in any case, an unequal status quo is often able to perpetuate itself only because of their support. From the standpoint of the favored party, a special virtue of norms of partiality is that in many cases, "the air of impersonality remains intact and successfully disguises what underlies the partiality norm, viz. an exercise of power" (p. 189). In this respect, norms of partiality are altogether different from PD norms and coordination norms.

Let's back up from the particular claims and notice that Ullmann-Margalit could be clearer about what, exactly, ensures that her categories of norms will have a stabilizing effect. She often speaks of social pressure, which leaves open the question: Would people follow norms if no one were watching? For PD norms, most of us would: You pay your taxes because it is the right thing to do, not because other people would think less of you if you didn't. The same is at least sometimes true of norms of partiality. If you are supposed to show deference to those of higher status (your boss, your teacher, a senator), you might well have internalized that practice, so that you would feel that you had done something wrong if you did not. (The term "impertinence" captures the idea.)

In *Explaining Social Behavior*, Jon Elster contends that "social norms operate through the emotions of shame in the norm violator and of contempt in the observer of the violation" (p. 355). This understanding helps to clarify much of Ullmann-Margalit's discussion, and it puts real pressure on her discussion of coordination norms. If the most sensible meeting place in New York City is Grand Central Station, and if you are the only one in a group of friends to show up at Penn Station, you might feel stupid, but shame would be a bit excessive. Ullmann-Margalit is aware of the problem and tries to distinguish between coordination norms and conventions, but I am not sure that the effort is successful.

For all three classes of norms, she is correct to emphasize the likely relevance of social pressure, but over the last decades, we have learned a great deal about that ambiguous idea. For example, Elster emphasizes ostracism, avoidance, and (perhaps more important) perceived contempt, and Ernst Fehr explores "altruistic punishment" (as when norm-enforcers take action at their own expense) and the anticipation of such punishment by would-be norm violators. We continue to learn about the immense power of shame, in the face of one's own norm violations, even when no such pressure is likely to be brought to bear. And of course legal norms, which Ullmann-Margalit sometimes seems to conflate with social norms, have an enforcement machinery of their own, even if they grow out of or codify social norms.

As her title suggests, Ullmann-Margalit is concerned with the emergence of norms, not only with their functions. The whole idea of "rational reconstruction" is designed to see how norms might plausibly have come about. (In this connection, she offers not the first but perhaps the clearest game-theoretic reading of Hobbes, in accordance with which government could emerge from a generalized PD-structured problem faced by humanity in the state of nature.) Those who emphasize rational reconstruction think that it is valuable to describe "the essential features of situations in which *such* an event *could* occur: it is a story of how something could happen – and when human actions are concerned, of what is the rationale of its happening that way – not of what did actually take place" (p. 1; emphasis in original). I might be missing something here, but wouldn't it be simpler and better to dispense with any kind of causal claims, even hypothetical ones, thus jettisoning an argument about the emergence of norms, and to speak instead of their functions?

A possible response is that if we are speaking of PD norms and coordination norms – and perhaps norms of partiality as well – their functions might turn out to have something to do with their emergence or at least their stability over time. Perhaps spontaneous orders generate norms of this kind. Elinor Ostrom's work so suggests, and Ullmann-Margalit's own work on invisible-hand explanations bears on the possibility (and also raises doubts about it). But the basic claim raises a host of questions, and they are empirical rather than conceptual in nature.

Taken in conceptual terms, and as analysis of functions rather than emergence, Ullmann-Margalit's discussion of PD norms and coordination norms remains broadly convincing (and contains many refinements that I have not been able to capture here). But with respect to norms of partiality, there is much more to say. Sure, some social norms stabilize situations of inequality, but Ullmann-Margalit acknowledges her own

struggle to identify clear examples. Private property and rights of inheritance are protected by law, not only and perhaps not mostly by social norms. As a matter of logic, rights of inheritance have to flow through legal institutions, and in advanced nations, no stable social norms specifies the content of those rights. In any case, private property can easily be understood as a solution to a generalized PD problem, and if they are taken as part of private property, rights of inheritance can be seen in just the same way.

With respect to norms of partiality, there is a mismatch between Ullmann-Margalit's very brief list of concrete examples and her extended statement of the abstract problem. In practice, norms that stabilize unequal situations might well be harder to maintain than PD norms and coordination norms – unless those norms are truly perceived as in the interest of those who seem to be disadvantaged by them (if, for example, they improve absolute position). To make progress, it would be useful to have a not-short catalogue of possible norms of partiality, understood as norms (Elster's sense) that disfavored people actually accept, or at least act as if they accept (because of a fear of sanctions).

Here's a possible direction. After the appearance of Ullmann-Margalit's book, Elster, Amartya Sen, and others have explored the idea of "adaptive preferences," by which disadvantaged people end up preferring their disadvantageous circumstances, and do not rebel against them. If your culture is pervaded by inequality, and if prevailing norms support that equality, you might not question them, and you might even end up accepting them (in part to reduce cognitive dissonance). Describing the hierarchical nature of pre-Revolutionary America, Gordon Wood writes that those "in lowly stations ... developed what was called a 'down look,' and "knew their place and willingly walked while gentlefolk rode; and as yet they seldom expressed any burning desire to change places with their betters." In Wood's account, it is impossible to "comprehend the distinctiveness of that premodern world until we appreciate the extent to which many ordinary people still accepted their own lowliness." Here, then, is a concrete account of adaptive preferences and their relationship to norms of partiality.

Alternatively, you might end up silencing yourself, and hence decline to rebel, not because you accept your own status, but simply because of reputational and other sanctions associated with rebellion – producing what Timur Kuran calls "preference falsification," which can contribute to social stability. Consider forms of inequality on the basis of sex, sexual orientation, and disability (physical and mental). Both adaptive preferences and preference falsification have played large roles, and they help to maintain norms of partiality. Both empirically and conceptually, there is a lot more to do on this subject.

In her final essay on considerateness, published posthumously and over thirty years after *The Emergence of Norms*, Ullmann-Margalit turned to an intriguing set of norms, by which people contribute to the well-being of others at low cost to themselves. She urged, quite boldly, that "considerateness is the foundation upon which our relationships are to be organized in both the thin, anonymous context of the public space and the thick, intimate context of the family." Focused not on the emergence of norms but their consequences, she notes that while a lover might send "a bouquet of a hundred roses," families typically have smaller, more routinized gestures and "deals," which

reflect "their preferences and aversions, their different competencies and skills, their relative strengths, weaknesses, and vulnerabilities, as well as their fantasies, whims, and special needs" (p. 221). Ullmann-Margalit argues that in that context, it may be too much to aspire to justice, but a good family can certainly be fair.

With respect to the family, Ullmann-Margalit was focused both on mutual advantage and on partiality. She insisted that we cannot proceed "'with eyes wide shut' – namely, in an imagined original position, behind a veil of ignorance" (p. 344). On the contrary, "the fair family deal is adopted considerately and partially, 'with eyes wide open' – namely, with the family members sympathetically taking into account the full particularity of each, and in light of fine-grained comparisons of preferences between them." Note that in this sentence, the word "partially" is paired with "considerately"; she sees a form of partiality as connected or at least compatible with fairness. Ullmann-Margalit thus began to explore a puzzle on which she was uniquely positioned to make progress: the role of norms of considerateness, and their highly complex functions not only within families and other close-knit units, but also within society as a whole.