Review of Edna Ullman-Margalit's *The Emergence of Norms*. Oxford, England,
The Clarendon Press of Oxford University Press, 1977. Pp. ix. 206.
Adrian M.S. Piper

Ullmann-Margalit's main thesis is that certain familiar types of norms -
for example, of honor, of driving, and of private property - can be rationally
reconstructed as solutions that emerge from three problems in game theory:
(1) Prisoner's Dilemma-type situations; (2) Coordination situations; and (3)
Inequality (or partiality) situations respectively. I must admit at the outset
that I am not sure of having represented her main thesis accurately: In some
places she talks as though she means merely to observe that (a) these norms
can be *correlated* with their respective game-theoretical situations as solutions
to problems posed by those situations: "It is the specific connection, in the
form of a solution, between norms and certain paradigmatic interaction
situations which is offered here as an account of the generation of these
norms, rather than a detailed description of some mechanism which actually
brings them into existence" (p. 12; also see pp. vii, 9, 10, 84). Here the idea
seems to be that the very fact, if it is a fact, that these types of norms solve
their corresponding game-theoretic problems justifies the claim that the
former emerge from the latter. In other places she seems to mean more: That
(b) these situations provide *necessary conditions* for the generation of their
corresponding types of norms (p. 5); or that (c) these situations represent
complex patterns of behavior of which their corresponding types of norms are
consequences (pp. 8, 11, 60); or that (d) the *functional effect* of these norms is to
prevent degeneration into their corresponding game-theoretical worst case
scenarios (pp. 103, 178-180).

Interpretation (a), above, of Ullmann-Margalit's main thesis is at once the
most definitive and most opaque explanation of what she means to be doing,
for it charges her with the following tasks. First, she must show that these
types of norms do, in fact, solve their corresponding game-theoretical
problems. Second, she must show that these types of norms have some sort of
special connection to their corresponding situations such that the fact that
they provide solutions for these situations justifies our nonarbitrarily
regarding them as generated by or emerging from these situations. Third, this
special connection, whatever it is, must not rely for its plausibility on some
mechanism that actually brings their corresponding norms into existence.
These tasks place certain constraints on the interpretation of Ullmann-
Margalit's main thesis, and on her success in defending it.

Do these types of norms solve their corresponding game-theoretical
problems? Consider first Ullmann-Margalit's treatment of Prisoner's
Dilemma-type situations. The problem is a familiar one: Two mutually
isolated, rationally self-interested individuals, A and B, are each confronted

by a choice: To sacrifice or to pursue self-interest.[1] If A and B each pursues self-interest, both lose heavily; if A pursues it and B sacrifices it, A wins the jackpot while B loses even more heavily (and vice versa); if A and B each sacrifices self-interest, both cut their losses and do relatively well for themselves (the second-best outcome for each). So it would be most rational for each to choose to sacrifice self-interest. Examples of sacrificing self-interest would include adherence to norms of honor that enjoin refusing to betray a jail buddy, refusing to desert when engaged with fellow soldiers in fighting a war (what Ullmann-Margalit calls "the Mortarmen's Dilemma," pp. 30-37), keeping promises or telling the truth even when personally disadvantageous, contributing to maintenance of the community public library, etc. In all of these cases, the general conclusion is supposed to have the same form: It is ultimately more rational for all individuals to adhere to such norms under these circumstances, because if each pursues self-interest, all lose heavily.

We can agree, for the sake of argument, that this is the general *conclusion* for Prisoner's Dilemma-type situations. But Ullmann-Margalit wants to claim that norms of the sort mentioned above are solutions to these situations, that is, that A and B can solve the dilemma by sacrificing self-interest. This is doubtful. What makes the Prisoner's Dilemma a dilemma for individuals caught in it is that, given their self-interested motivations, each cannot try to maximize her self-interest without accomplishing just the opposite, and neither can try to sacrifice her self-interest without risking an even greater loss in case the other does not. The dilemma is not solved by A's and B's verbal agreement to sacrifice self-interest, since, as Hobbes knew, the same dilemma may resurface with respect to the status of this agreement. Since we do not know what shared social norms are already motivationally effective for A and B, the dilemma raised by their self-interested motivation may infect all of them. So long as A and B are stipulated to be wholly self-interested, there is no way out of the dilemma. Call this a *sealed dilemma*: If the Prisoner's

---

[1] Considerations of space preclude addressing deep and difficult questions about whether, if A and B each chooses to sacrifice immediate self-interest when the other does in order to obtain the second-best outcome, they should be understood as maximizing self-interest in some broader or more extended sense. For purposes of this discussion I shall abide by the convention of using "self-interested" to mean what Sen (in "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," *Philosophy and Public Affairs* 6 (1977), pp. 317-344) calls "selfish," by contrast with the broader, genuine rationality that recommends such a sacrifice. Since the contrast is preserved even when this brand of self-interest is taken to include interests of, not just in a self (cf. John Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), Sec. 22, p. 127), my use of the term "self-interest" will also cover this inclusive sense.

Dilemma obtains for two individuals, there is no strategic solution to that dilemma under the motivational stipulations that make it a dilemma for them in the first place.[2]

This problem can be dissolved by *deus ex machina* measures of various kinds. This seems to be the way Ullmann-Margalit conceives her task: "The problem," she claims, "is to find some means by which to deprive the [self-interested] strategy of its tempting force; that is, to change the situation - or the degrees of desirability of the outcomes - in such a way as to cancel the dominance of the [self-interested] choice over the [honorable] choice" (pp. 31-32). In particular, she proposes that stipulating the parties to be motivated by internalized norms of honorable behavior will have this effect (pp. 36-41).[3] That changing A's and B's motivations - through coercion or the inculcation of altruistic norms - will dissolve the Prisoner's Dilemma in which they are trapped is undeniable. If A and B are motivated by self-interest under these circumstances, then they are trapped in a sealed dilemma, and if they are not, then they are not; this much goes without saying. This is why the *conclusion* for Prisoner's Dilemma-type cases is ordinarily that genuinely rational individuals are not motivated exclusively by considerations of self-interest. But we are driven to this conclusion because we recognize that there is no *solution* to the Prisoner's Dilemma if they are, and no dilemma if they are not. Ullmann-Margalit's solution was supposed to exhibit a special relation to this sealed dilemma that justified our regarding that solution as having *emerged out of* it, and not just imposed upon it from without. It was supposed to *solve* the problem, not just *dis*solve it. But if the only solution to the Prisoner's Dilemma requires changing the motivational assumptions that structure it,

---

[2] Notice that the dilemma remains sealed even if we suppose A and B to conclude, as we have, that genuine rationality consists in sacrificing self-interest under these circumstances. For each may realize this, without trusting the other's conversion to genuine rationality. In this case, each will be motivated by self-interest to refuse to sacrifice self-interest.

[3] Another *deus ex machina* measure would be to introduce into the structure of the situation a device that would enable each to ascertain independently the other's choice and choose accordingly, as does David Gauthier's stipulations of motivational transparency and translucence in Chapter VI of his *Morals by Agreement* (New York, N.Y.: Clarendon Press of Oxford University, 1986), pp. 174-178. However, this would be a "mechanism which actually brings [these norms] into existence," and we have already seen that Ullmann-Margalit explicitly rejects adverting to any such mechanism. On the other hand, since she also thinks that a high risk of physical harm or discipline backed by punishment are alternative solutions (pp. 31-36), it is unclear, in the absence of any such mechanism, why we should suppose Prisoner's Dilemma-type situations to generate internalized norms rather than either of these alternatives.

and if no such change can be effected from within such a sealed dilemma, then no such solution can be shown to emerge from it in any intuitive sense. And since Ullmann-Margalit has exhibited no resources *within the structure of the sealed dilemma as originally described* for bootstrapping A's and B's motivation from self-interested into genuinely rational, she has provided no reason to view the norms she discusses as having emerged from this type of dilemma.

Ullmann-Margalit's treatment of coordination situations is subject to the same objection. Here she relies on the apparatus developed by Schelling and Lewis[4] to show that norms can be generated by coordination situations without the aid of explicit agreements among the parties involved (p. 76). Coordination situations are those in which two or more parties must each manage to choose the same alternative that the other chooses, whichever that is, in order for each to maximize self-interest. For example, in order for you and me to resume our unexpectedly disconnected phone conversation, either I must call back and you must wait, or vice versa; in order for each of us to drive without undue risk of auto accidents to any of us, all of us must drive either on the right or on the left. In all such cases, the success of each party requires choosing that alternative she expects the other one to choose among a set of mutually indifferent ones, knowing that the other is choosing in accordance with expectations of what she is likely to choose. Here Ullmann-Margalit means to argue that "(1) In a recurrent co-ordination problem a successful solution, *once arrived at and thence repeated*, becomes a norm. (2) In certain novel co-ordination problems a solution is likely to be dictated by a norm issued specifically for that purpose by some authority" (p. 83; my emphasis). The kind of "solution" Ullmann-Margalit envisions for coordination problems of the first kind would be that, for example, you and I reach an understanding that when our telephone conversation is unexpectedly disconnected, I will call back and you will wait; for problems of the second kind, that all of us agree to abide by the traffic law that requires us to drive on the right.

But this is to revert once more to a *deus ex machina* intervention. Again we can agree that abiding by either of these norms will *dis*solve the coordination problem. For if they are available resources for the participants to begin with, there is no problem: If there is a traffic law that requires driving on the right, then the question of which side of the road to drive on does not arise. Ordinarily we think that something *solves* a coordination problem by showing

---

[4] Specifically, Thomas Schelling, *The Strategy of Conflict* (Cambridge, Mass.: Harvard University Press, 1980), Chapter I.3, esp. pp. 54-58; David Lewis, *Convention* (Cambridge, Mass.: Harvard University Press, 1969), Chapters I and II.

us *how* to coordinate our choices, not merely by coordinating them. A solution to a coordination problem enables me to decide which alternative I should choose, such that, for example, I can figure out whether or not you will wait for my call, or such that we each can ascertain on which side of the road to drive. This is the lacuna that makes coordination a *problem* for us, when it is a problem, and that is required to be filled in order to solve it.[5] But the above-mentioned norms do not fill this lacuna. They settle in advance the question of which alternative I should choose without enabling me to figure out which alternative I should choose. Since they do not solve the coordination problem, interpretation (a) of Ullmann-Margalit's main thesis would thus appear not to apply to so-called coordination norms. For these do not show us how to arrive at a coordination, and this is what a solution to a coordination problem needs to do.

Such a solution is, however, provided by the apparatus Ullmann-Margalit adopts from Schelling and Lewis. In particular, she acknowledges - inconsistently, I think - that "co-ordination problems are, very generally, solved through *salience*: one of the co-ordination equilibria might appear conspicuous to the people involved, owing to some specific feature it possesses, and might hence serve as a focal point for the convergence of their choice of actions" (pp. 83-84). Here the idea is that if we already know something about the habits and characteristics of the other participants, we may be able to form an expectation that one particular alternative may seem more vivid or desirable to them than others, and choose accordingly. For example, if you know that I have arthritic hands, you may predict that I will wait for you to call back rather than calling back myself, and so choose to call back. Having once called back, you may then get into the habit of calling me back whenever we are disconnected. Your calling me back when we are disconnected thereby attains the status of a norm that coordinates our behavior under these circumstances. But what solves the problem of coordination between us is not the norm that you call me back when we are disconnected. Rather, it is your perception that my hands are arthritic, my

---

[5] Schelling and Lewis concur with this characterization of a solution to a coordination problem, and both defend some shared perception that generates mutual expectations as the prime candidate. Here is Schelling: "What is necessary is to coordinate predictions, to read the same message in the common situation. . . . [Two individuals] must 'mutually recognize' some unique signal that coordinates their expectations of each other (ibid., p. 54)… A prime characteristic of most of these 'solutions' to the problems, that is, of the clues or coordinators or focal points, is some kind of prominence or conspicuousness" (ibid., p. 57). Lewis elaborates the thesis that we rely on a system of mutually concordant expectations to achieve coordination: Agreement is one means of producing those expectations, salience another (ibid., Chapter 1.3).

expectation that you will perceive this, your expectation that I expect you to perceive this, and so on.

Even though salience, and not the norms generated by it, solves such coordination problems, Ullmann-Margalit nevertheless has justification for claiming in this case that norms emerge from coordination problems. In particular, she could say, they emerge from the solutions to coordination problems, that is, from salience. Salience provides the special connection of norms to their corresponding coordination situations that justifies our nonarbitrarily regarding them as generated by or emerging from these situations. Unfortunately, this special connection relies for its plausibility on precisely that mechanism, that is, salience, that actually brings their corresponding norms into existence. And Ullmann-Margalit reiterates that she has no interest in any such mechanism: "Since our prime concern at present is not with the precise way a solution was arrived at in the first place, but with what character it assumes after it has been achieved, repeated, and established, [special cases of salience] will not here be further dwelt upon" (p. 84). She thereby undercuts the only resource she offers to buttress her claim to be giving necessary conditions for the generation of norms (interpretation (b) of her main thesis), or that norms are consequences of these coordination situations (c).

Do these norms function to prevent deterioration of the coordination situation into the worst-case scenario, that is, in which all participants are paralyzed by the inability to decide among the alternatives (interpretation (d) of Ullmann-Margalit's main thesis)? This depends on what norms are already in place, and Ullmann-Margalit does not tell us this. If there are no prior stable norms to coordinate any of our expectations or linguistic behavior, then these norms, and the salient characteristics that generate them, will not help us; for the same coordination problem of establishing these norms will arise for establishing the salient characteristics on which they are based.[6] In this case, the coordination problem is a sealed dilemma, for there is no solution to the problem that does not require altering at least some of the motivational stipulations that make it a problem in the first place. In particular, we must assume that at least some well-founded mutual expectations already have been established.

So suppose we make this assumption. In this case, my worry concerns the application of interpretation (d) to Prisoner's Dilemma situations as well as coordination ones: We can agree that without such norms the situations would deteriorate into their respective worst-case scenarios. But it is unclear

---

[6] This point is discussed at greater length in my "Utility, Publicity, and Manipulation," *Ethics* 88 (1978), pp. 189-206.

why we need to suppose ourselves ever to have been without such norms, and how or why we should rationally reconstruct them as having emerged from such problematic situations. By aiming to defend the thesis that shared social norms can be explained as emergent from Prisoner's Dilemma or coordination situations, Ullmann-Margalit has arrogated to herself the important and challenging task of defending the widely assumed game-theoretic premise of the motivational priority of self-interest over those norms in the structure of the self. But I sought in vain throughout this discussion for attention to this task. The *de facto* existence of such norms, plus the impossibility of generating them from sealed dilemmas of either variety, strongly suggests that these dilemmas can be neither necessary conditions of such norms (interpretation (b)), nor result in them (interpretation (c)).

Finally, I want to discuss briefly Ullmann-Margalit's treatment of inequality or partiality situations. Inequality situations are those in which the benefits of the situation are unequally distributed between the two parties in them, and in which the disadvantaged party is motivated to improve her position whereas the advantaged party is motivated to defend the *status quo* that accords her advantaged status. Since inequality situations are by hypothesis interactional, the strategic alternatives open to the disadvantaged party for improving her position require the cooperation of the advantaged party, and therefore that the advantaged party find such cooperation optimal for her: "This inducement will be achieved once one succeeds in affecting the other party's expectations concerning one's own behavior. And one will be able to influence the other's expectations through visibly and persuasively constraining one's own behavior" (p. 165). I may achieve this if I convince you that, regardless of any attempts of yours to placate me, my disadvantaged position is so unacceptable to me that I am prepared to abandon the *status quo*, by destroying both of our benefit allocations, in order to escape it. I thereby demonstrate to you that achieving equality with you is more important to me than protecting the inferior benefits I have. Then you have reason to abandon the *status quo* yourself by merely abdicating your advantage, in order to contain the damage and avert the disaster of "equality in misery." Such is the reasoning behind much black South African support for international economic sanctions. How might the advantaged party defend the *status quo* against these alternatives? Among six possible means Ullmann-Margalit evaluates, she finds fortification by norms the most effective: "[O]nce it is in some sense normatively required that the *status quo* endure, the nature of the possible calculations and considerations of deviance functionally changes: it is no longer evaluated only in terms of being 'costly' or 'risky', but as being *wrong* or *subversive*" (p. 170). Under these circumstances, the disadvantaged can be discouraged from abandoning the

status quo not only by punishments, but by moral disapproval. The norms associated with private property would be an example.

By contrast with Prisoner's Dilemma and coordination situations, inequality situations are not structured to presuppose an allocation of roughly equal resources to each participant. They therefore offer a type of case in which it is easier to see how partiality norms might actually be strategic solutions to the problem of how to maintain the *status quo* of inequality, how they might emerge from within the structure of inequality situations, and how they might function to prevent destruction of the *status quo*. Basically, the strategy would be to use some of one's advantage to make the disadvantaged feel so guilty about staking a claim to one's benefits that improving their position is no longer a preferred option. One would think that Ullmann-Margalit would seize the chance to apply the terms of her analysis to this case. But she does not. She does assert that "[partiality norms] might be conceived of as a sophisticated tool of coercion, used by the favoured party in a *status quo* of inequality to promote its interest in the maintenance of this *status quo*. It will be considered sophisticated to the extent that the air of impersonality remains intact and successfully disguises what really underlies the partiality norms, viz. an exercise of power" (p. 189). But she denies the possibility that this might be to any degree a conscious agenda among the advantaged (pp. 180-183), and suggests instead that such strategic behavior on the part of the advantaged is merely predetermined by one's social role - as though this settled the question of whether such behavior is undertaken consciously or not, or with some degree of complicit understanding of the purposes it serves. However, this question is not settled, nor is any competing account advanced to explain how these norms might emerge from inequality situations, if not from purposeful and self-interested behavior on the part of those concerned to preserve them. And without such an account, we remain in the dark as to how these norms emerge or result from inequality situations, and how they fit into the parties' strategic thinking as a solution.

There is much in this book I have not mentioned. It is an extremely thorough study of the cases she targets, and the author's discussion of partiality norms is particularly original. She is tremendously erudite, and there are suggestive digressions on many topics, from MacPherson on Locke to Dahrendorf to functionalism to generalization in ethics to envy and rationality to corporate cartels to equilibrium versus stability to theoretical versus deontic expectations, and much more. However, most of these are tangential to Ullmann-Margalit's stated project, which, as I have tried to show, she largely neglects. There are in addition many confusions - about the distinctions between norms and conventions, norms and conformity to them, norms and regularities and salience, norms and the social choices resulting

from them, and so forth. I have found this book exceedingly difficult to read. The prose has a stilted, labored quality to it, and too much space is devoted to elementary exposition and compare-contrast exercises. As though it were the proverbial broth to which too many cooks had contributed, the thread of the argument is often so obscured by extraneous issues the author feels obliged to address that she herself must recapitulate the main point sometimes two or three times in a section. Ullmann-Margalit covers so many topics of interest and significance that one hopes she will focus her considerable resources on each of them in depth in future works.