Foundations of institutional reality, by Andrei Marmor, New York, Oxford University Press, 2023

Samuele Chilovi (Institute of Philosophy, CSIC)

Published in *Jurisprudence*, 15:4, 587-593
https://doi.org/10.1080/20403313.2024.2380181
Penultimate draft: Please refer to the published version

Andrei Marmor's rich and penetrating new book, *Foundations of Institutional Reality* (hereinafter *FIR*), outlines a detailed and comprehensive account of institutional reality, of its nature and grounds.[1] This account has a reductive aim, i.e. to show how complex collective social practices – such as games, languages, money and law – are built from interactions between large collections of individual actions, attitudes and mental states.

Along the way, Marmor takes up a number of diverse but interconnected topics. In chapter 1 he introduces some of the book's key concepts (such as the notions of an institutional fact, a social practice and a rule), and singles out his object of interest as the fragment of social reality he calls 'institutional'. In chapter 2 he lays out some assumptions regarding the metaphysical tools used in the book (such as grounding and reduction), and clarifies his project's reductive ambitions. He distinguishes two notions of reduction – identity reduction and grounding reduction – and sets himself the goal of providing a reduction of the latter kind.

In chapters 3 and 4, Marmor develops the central elements of his account. In chapter 4 he argues that institutional facts are partly, though not fully, grounded in social rules. In so doing, he takes issue with significant parts of Searle's theory.[2] In chapter 3 he proposes a reduction of social rules to behavioural and psychological elements which builds on, but significantly revises, Hart's influential view on this matter.[3]

In chapter 5 Marmor derives some epistemic consequences regarding the possibility of error about certain aspects of social practices and rules. In chapter 6 he rejects the supposition – attributed to Dworkin[4] – that an account of social practices must

[1] Andrei Marmor, *Foundations of Institutional Reality* (OUP 2023).

[2] John Searle, *The Construction of Social Reality* (Free Press 1995).

[3] HLA Hart, *The Concept of Law* (OUP 1961).

[4] Ronald Dworkin, *Law's Empire* (HUP 1986).

meet the 'internal rationalization constraint', i.e. must 'offer the kind of explanation that would rationalize the practice for its participants, make it intelligible to them why they engage in it, what the point of it is, for them'.[5] Finally, in chapter 7, Marmor delineates the place of normative powers within the social structure depicted by the book, and argues that power-conferring rules play a fundamental role in the construction of certain parts of social reality, such as legal systems.

In this review, I shall focus on the reduction of social rules developed in chapter 3. This is a central part of Marmor's account, critical to his overall aim of explaining institutional reality in ultimately non-normative and individualistic terms. He motivates *FIR*'s distinctive view as a solution to a circularity problem that allegedly affects Hart's theory. Here I shall argue, first, that Marmor's characterisation of this problem is wanting; second, that if it gives rise to a genuine concern at all, it is better addressed in other ways; and third, that there are other problems to which both Hart's and *FIR*'s accounts are vulnerable. Before we get there, let me say a few words about how Marmor identifies *FIR*'s object of interest.

As mentioned before, *FIR*'s target of analysis is institutional facts, where these are characterised as a specific kind of social fact. Whereas social facts are any facts that ontologically depend on (i.e. modally require) certain interactions between humans,[6] institutional facts are those social facts that depend on (in a stronger, to-be-articulated sense) the existence of social rules. Given these stipulations, the thesis that institutional facts depend on rules isn't substantive or controversial, but rather merely serves to demarcate the author's object of inquiry. This, however, leaves open what sorts of facts and practices – if any – count as institutional in the defined sense. Marmor assumes – plausibly, I think – that this class is non-empty, citing games, languages, money and law as significant exemplars of it.

Once these premises are in place, the metaphysical questions that *FIR* tackles concern how institutional facts are explained by rules, and how social rules may be reduced to more basic elements. In what follows, I focus on the latter.

---

[5] Marmor (n 1) 105.

[6] ibid 1. Marmor is open to there being social facts that arise from interactions among non-human animals, but sets them aside for the purposes of his book.

The starting point for *FIR*'s account is Hart's practice theory of rules,[7] which Marmor formulates in the following way.[8] A social rule R exists in a population S if and only if:

(1) Most members of S regularly conform in their behaviour to the content of R.
(2) Most members of S accept R as a rule, which means that
    (a) for most members of S the existence of R constitutes a reason for action in accordance with R; and
    (b) members of S tend to employ R and refer to it as grounds for exerting pressure on other members to conform to R, and as grounds for criticising deviations from conformity to R.

I will set aside the question of whether this rendering of Hart's view is accurate. My primary focus will be on the problems that *FIR* finds in it and the revisions it proposes to solve them. Before we examine these, let me note that although Marmor ultimately rejects Hart's theory in its stated form, he shares two of its central tenets. First, he endorses its ambition to explain a normative notion – that of a social rule – in purely non-normative – behavioural and psychological – terms. Second, both Marmor and Hart's accounts are designed to vindicate a form of 'methodological individualism', that is, explain a collective social phenomenon by reference to facts about individuals. So both of these theories are broadly naturalistic and individualistic.

That said, the problem that motivates Marmor to depart from Hart's theory is that it provides a viciously circular definition of what it is to be a social rule, and for this reason fails to tell us what a social rule is. As he puts it, on Hart's theory 'we have a [social rule] when, inter alia, people regard the *existence of the rule* as a reason for action'.[9] Since this theory 'lacks a noncircular account of what it is that people in S take to be their reason for action',[10] it must be rejected.

To remedy this problem, Marmor recommends a twofold revision: the account should be neither circular, nor a definition. To accomplish the first task, the account's

[7] Hart (n 3).

[8] Marmor (n 1) 11.

[9] ibid 12.

[10] ibid 41.

conditions are amended, and to fulfil the second, they are interpreted so that they are taken to provide a grounding reduction, rather than an identity reduction, of social rules. Let's take up these amendments in turn.

On Marmor's account, what explains the existence of a social rule R in a population S are the following facts:[11]

(1*) Most members of S regularly conform in their behaviour to R's prescriptive content.

(2*) Most members of S commonly know that R's content is collectively intended in S and regard this fact as giving them a reason to conform and reason to exert pressure on others in S to conform as well as criticise them when they don't.

The first clause merely restates the behavioural element of Hart's account, requiring for the existence of a social rule that most members of the group comply with its content. The psychological clause, in contrast, makes several changes, the main one being that the mental attitudes required for the rule's existence take as their object not the rule itself, but rather a collective intention. Consider for example a social rule R_i, in S_i at time t_i, according to which every F ought to φ in circumstances C. What explains the existence of R_i, according to Marmor, is that most members of S_i know that *it is collectively intended* in S_i that every F ought to φ in circumstances C, and that they take *this* as a reason for action as well as grounds for exerting pressure on others.

However, revising the account's psychological element by placing collective intentions at its core is only seen as part of the solution. For, Marmor says, although this modification 'goes some way toward resolving the circularity problem', we need one more step. That is, to move away from conceiving of the account as giving an *identity* reduction, and regard it instead as a *grounding* account of social rules, with facts about conduct and collective intentions furnishing the metaphysical grounds for the existence of the rule.[12] Presumably, the idea here is that interpreting the relation between the

[11] ibid 53.

[12] ibid 52. There is a large and growing literature on the notion of metaphysical grounding. For some earlier work on this, see Jonathan Schaffer, 'On What Grounds What' in David Manley, David J Chalmers and Ryan Wasserman (eds), *Metametaphysics: New Essays on the Foundations of Ontology* (OUP 2009); Gideon Rosen, 'Metaphysical Dependence: Grounding and Reduction' in Bob Hale and Aviv Hoffmann

reduced entity and its reduction base as one of grounding would ensure that circularity is avoided, thanks to the fact that the grounding relation is irreflexive and thus incompatible with identity.[13]

Several elements of Marmor's solution deserve attention. First, it is unclear why both parts – removing the circularity from the clauses *and* interpreting the account as a grounding explanation – are needed. If the account is no longer circular because collective intentions take the place of rules in the reduction base, then why must it also be incompatible with identity? Not only is the latter unnecessary for solving the circularity problem, it also appears to preclude an alternative solution. For note that if the reduction were compatible with identity, then it would *not* necessarily be imperilled by the fact that the same entity appears on both sides of the reduction.

At any rate, the key element of the solution lies in its revision of the account's clauses. The vicious circularity supposedly sprang from the fact that the notion being accounted for – a social rule – also appears on the side of what should account for it. Do the new clauses avoid this? At the surface level, the new clauses do mention the social rule they are the grounds of. As per (1*), the existence of R is grounded in behaviour that complies with the content *of R*, and, as per (2*), the existence of R is grounded in shared knowledge that the content *of R* is collectively accepted, among other things. If the standard by which we judge the presence of vicious circularity were met merely by mentioning the social rule on both sides of the account, then this account would fail it too.

But it would be unfair to judge the issue by this standard. To see this, suppose that we were trying to reduce the property of being a bachelor to the property of being an unmarried man. Then we could say that for any human being H, H is a bachelor iff (and because) H is an unmarried man. And this isn't relevantly different from what we do when, in trying to reduce the notion of a social rule to some other features (say, collective acceptance), we claim that for any rule R and population S, R is a social rule in S iff, and

(eds), *Modality: Metaphysics, Logic, and Epistemology* (OUP 2010); Karen Bennett, 'By Our Bootstraps' (2011) 25 Philosophical Perspectives 27; Fabrice Correia and Benjamin Schnieder (eds), *Metaphysical Grounding: Understanding the Structure of Reality* (CUP 2012).

[13] Though for a challenge to the irreflexivity of grounding, see Carrie Jenkins, 'Is Metaphysical Dependence Irreflexive?' (2011) 94 The Monist 267.

because, *R* is collectively accepted in S. The fact that we're explaining R's being a social rule in terms of *R's* satisfying some conditions doesn't make the account circular.

What then – if anything – is the source of the circularity in Hart's account? At one point, Marmor describes the circularity as resulting from the fact that for R to be a rule, *the existence of R* needs to be regarded as a reason for action.[14] But this is just a convoluted way of saying that for R to be a rule, R needs to satisfy a certain substantive condition – being regarded as a reason for action – and there is certainly nothing circular about defining the notion of a social rule in terms of being regarded as a reason for action.[15]

A more plausible reason for taking Hart's account (as formulated by Marmor) to be circular might come from the idea that for R to be a rule, R needs to be accepted *as a rule*. But if this is the source of the circularity, it is easily fixed. Remember that this condition is unpacked via two clauses: 2(a) For most members of S the existence of R constitutes a reason for action in accordance with R; and: 2(b) Members of S tend to employ R and refer to it as grounds for exerting pressure on other members to conform to R, and as grounds for criticising deviations from conformity to R.

Now, these clauses are ambiguous between two interpretations, only one of which is problematic. The problematic interpretation is that for R to be a social rule, most members of S must regard *the fact that R is a social rule* as a reason for action, and use *this fact* as grounds for exerting pressure (etc.).[16] But a more charitable interpretation of Hart's account is available, namely that for R to be a social rule, most members of S regard *R* as a reason for action, and use *R* as grounds for exerting pressure (etc.). And this, as we've seen, is completely unproblematic.

The fix proposed by Marmor – of letting the fact that *the content of R is collectively intended* be the object of the population's psychological states and attitudes

---

[14] Marmor (n 1) 12.

[15] Notice, further, that if the mere presence of reference to the rule were enough to generate circularity, then the first clause of the account – that for R to be a rule, R needs to be generally complied with – would equally spell trouble.

[16] This highly implausible reading of the clauses appears to be suggested by Marmor when he says: 'To say that there is a social rule when people believe that *there is a social rule* and take that [i.e. that *there is a social rule*] to be a reason for their actions is not a good definition' (52, emphasis added).

(such that this content is known and regarded as reason-giving) – is therefore quite unmotivated. For if the circularity were rooted in R's appearance in the reduction base, then the circularity wouldn't be eliminated. And if it derived from the population's acceptance of R *as a rule*, then it could be avoided through a charitable interpretation of Hart's account.

As a consequence, I take Marmor's and Hart's accounts to be on a par in terms of their ability to avoid circularity. Let me now conclude by making a few points about their substance. First off, as we've seen, Hart's and Marmor's views both appeal to a combination of a behavioural element and a psychological element, manifested in the conduct and mental states of most members of the relevant population. As to conduct, both accounts agree that for a rule to exist, most members of the population must conform in their behaviour to its content. As to psychology, both accounts require that some mental attitude or state be held by most members of the population.

This is problematic, for various reasons. For one, it is unclear why there should be a *universal* cutoff point, such that rules exist only if the relevant individuals' attitudes and behaviours surpass it. Rather, it seems more plausible to say that the amount of attitudes that are necessary (and sufficient) for the existence of a rule should be able to vary, depending on which rule and population we are considering. Moreover, even if a universal and sharp cutoff point were to exist, it seems arbitrary that it should amount to *most* members of the population performing the right conduct and holding the right attitude. Why not more, or less? Rather, it would be much more plausible for 'most' to be substituted by 'enough' in these accounts. To say that *enough* members should display the required attitudes and behaviour is suitably flexible, for it doesn't specify any exact proportion of the population.

Moreover, the conduct element has been subjected to compelling criticism on its own. Brennan et al. challenge it by imagining a community accepting a rule that one must not urinate in public swimming pools.[17] Members of this community judge that one must not do so, they take this as a reason not to, they are disposed to criticise those who do, etc. Yet it turns out that most – perhaps all – violate this norm. They do so secretly, of course. When they urinate in pools, they feel guilty about it, and when on sporadic

[17] Geoffrey Brennan; Lina Eriksson; Robert E. Goodin; & Nicholas Southwood. Explaining Norms. (Oxford: Oxford University Press UK 2013).

occasions they catch someone else doing it, they are outraged. In such a scenario, it seems intuitive that there is a social rule, despite most of the population violating it, meaning that compliance with a rule cannot in general be a necessary condition for its existence.

Second, Marmor's account takes the notion of a reason used in both his and (his rendering of) Hart's account to be motivating rather than normative. A normative reason for action (or belief) is a consideration or fact that counts in favour or against doing (or believing) something. A motivating reason, in contrast, is a consideration or fact that the agent *takes* to count in favour or against doing (or believing) something. Normative reasons are reasons *for* (or against) doing something, whereas motivating reasons are the reasons *for which* one does (or does not do) something.

Marmor claims to be following 'Hart in talking about reasons for action in the causal-motivating sense'.[18] He states:

> Whether a rule, or anything else is, normatively speaking, a reason for action or not is beyond the concerns of the metaphysics of sociality, quite generally. We have to be careful not to turn a metaphysical account of rules into a normative one. Thus, from now on, talk about reasons for action is talk about motivating reasons, not normative ones.[19]

A few observations about this are in order. Setting aside whether Marmor is right to view the notion of reason employed by Hart as motivating, the main question is which of these two notions is more suitable for an account like Hart's (or Marmor's). Marmor claims that whether a rule is, normatively speaking, a reason is beyond the concerns of the metaphysics of social rules. This just strikes me as implausible. It would be extremely relevant to the metaphysics of social rules to know whether they provide normative reasons or not. To know whether they do is part of what one legitimately expects from an account of the nature of such rules. So an account of social rules which held that social rules provide normative reasons would not thereby cease to be a metaphysical account!

[18] Marmor (n 1) 10.

[19] ibid 10-11.

Perhaps what Marmor means, then, is that a metaphysical account of social rules *of the kind that he and Hart are giving* should deny that social rules provide normative reasons. One hypothesis is that this commitment stems from their adherence to legal positivism – roughly, the view that legal facts are fully grounded in social, non-moral facts – together with the role that social rules play in their positivist accounts of law.[20] But there is no inconsistency between taking legal facts to be grounded in social rules and thinking that social rules provide normative – even robustly moral – reasons for action, as long as the fact that they do isn't used to ground the legal facts themselves.[21]

Another hypothesis is that this commitment runs downstream from the naturalistic nature of their view, which seeks to explain a normative entity (a rule) in terms of naturalistic facts (mainly, about behaviour and mental states). But even this suggestion seems to be misplaced. For it seems perfectly consistent to hold that social rules provide reasons for action, even though they are fully grounded in descriptive natural facts. This combination of views, in fact, should look especially attractive to someone who, while naturalistically inclined, doesn't wish to deny that social rules have some sort of normative force.

In contrast, something that would create a *prima facie* tension with a naturalist explanation of social rules would be to claim that the fact that R provides normative reasons *grounds* the fact that R is a social rule. But neither Hart's nor Marmor's account would say anything of the sort, even if the notion of reason they invoked were normative. In both cases, the fact that R provides a reason is embedded in a psychological state or attitude. For Hart, R is a social rule in S only if S's members *regard* R as providing reasons for action; while for Marmor, R is a social rule in S only if S's members *regard* the fact that R is collectively intended as providing reasons for action. In either case, what does

[20] For the use of metaphysical grounding in defining legal positivism, see Samuele Chilovi, 'Grounding-Based Formulations of Legal Positivism' (2020) 177 Philosophical Studies 3283; Samuele Chilovi and George Pavlakos, 'Law-Determination as Grounding: A Common Grounding Framework for Jurisprudence' (2019) 25 Legal Theory 53; Samuele Chilovi and George Pavlakos, 'The Explanatory Demands of Grounding in Law' (2022) 103 Pacific Philosophical Quarterly 900; Samuele Chilovi and Daniel Wodak, 'On the (In)significance of Hume's Law' (2022) 179 Philosophical Studies 633.

[21] See Rosen (n 12) for a formulation of naturalism.

the explanatory work is not the putative fact that R provides a normative reason, but rather the fact that it (or its being collectively intended) is taken to provide one.

Notice also that taking the notion of a reason in Hart's and Marmor's formulations to be motivating would make these formulations convoluted, redundant, and ultimately implausible. Hart's condition (2a) would then say that for R to be a social rule in S, members of S should regard R as among the motivating reasons they have for doing as R says. But it seems entirely irrelevant to Hart's account that S should have any *beliefs about their own motivations*, i.e. about the psychological explanation of why they behave as they do. The group's members, on (a plausible version of) Hart's account, need to *treat R as giving them a reason* (whether or not R does in fact give them any such reason); they need not engage in any speculation about their own psychology or what moves them to action. Similarly, it seems very implausible to require of the population that they believe that *what motivates them to action* is the fact that R is collectively intended. Again, any speculative hypotheses possibly accepted by these agents about their own motivations would seem irrelevant even by the lights of Marmor's view.

In conclusion, the account of social rules presented in *FIR* is simple, elegant, in some ways plausible, and explanatorily powerful. It is certainly a serious contender – along with Hart's and others' accounts – for explaining how rules arise from collective patterns of attitudes and/or behaviour. While I have tried to show that it suffers from some serious problems, and that it is not as strongly motivated as one might have thought, it remains to be seen how it might be able to solve these problems and find alternative sources of support.[22]

[22] Acknowledgments. This work has been supported by the Ramón y Cajal fellowship 'The Metaphysical Structure of Normative Explanation' (ref.: RYC2021-032064-I) at the Institute of Philosophy (IFS) of the Spanish National Research Council (CSIC), funded by the Spanish Government (MCIN/AEI/ 10.13039/501100011033) and by the EU (NextGenerationEU/PRTR). It has also been supported by the project 'Razón Pública Global' (ref.: PID2020-115041GB-I00) at Pompeu Fabra University, funded by the Spanish Government (MCIN/AEI/10.13039/501100011033).