Author Query Form

Journal: *Oxford Journal of Legal Studies*
Article Doi: 10.1093/ojls/gqy039
Article Title: Social Norms and the Internal Point of View: An Elaboration of Hart's Genealogy of Law
First Author: Philip Pettit
Corr. Author: Philip Pettit

AUTHOR QUERIES – TO BE ANSWERED BY THE CORRESPONDING AUTHOR

The following queries have arisen during the typesetting of your manuscript. Please click on each query number and respond by indicating the change required within the text of the article. If no change is needed please add a note saying "No change."

AQ1: Meaning clear?

AQ2: Title of Green's contribution? 'Introduction', as mentioned in n 1.

AQ3: This is the second section. Please clarify

AQ4: Journal title changed - OK? Fine

AQ5: What is it?

AQ6: What is (Pettit 2018a)? Please give full ref.

AQ7: Please check that all names have been spelled correctly and appear in the correct order. Please also check that all initials are present. Please check that the author surnames (family name) have been correctly identified by a pink background. If this is incorrect, please identify the full surname of the relevant authors. Occasionally, the distinction between surnames and forenames can be ambiguous, and this is to ensure that the authors' full surnames and forenames are tagged correctly, for accurate indexing online. Please also check all author affiliations.

NOT FOR PUBLIC RELEASE

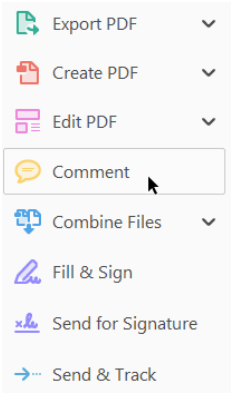
MAKING CORRECTIONS TO YOUR PROOF

These instructions show you how to mark changes or add notes to your proofs using Adobe Acrobat Professional versions 7 and onwards, or Adobe Reader DC. To check what version you are using go to **Help** then **About**. The latest version of Adobe Reader is available for free from get.adobe.com/reader.

DISPLAYING THE TOOLBARS

Adobe Reader DC

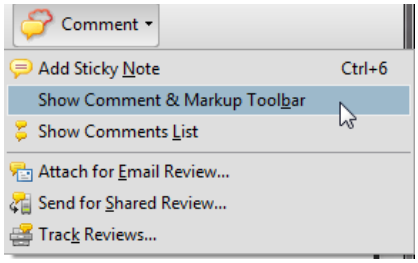
In Adobe Reader DC, the Comment toolbar can be found by clicking 'Comment' in the menu on the right-hand side of the page (shown below).



The toolbar shown below will then display along the top.

Acrobat Professional 7, 8, and 9

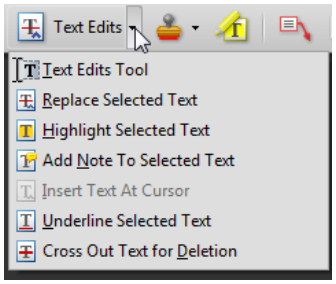
In Adobe Professional, the Comment toolbar can be found by clicking 'Comment(s)' in the top toolbar, and then clicking 'Show Comment & Markup Toolbar' (shown below).



The toolbar shown below will then be displayed along the top.

USING TEXT EDITS AND COMMENTS IN ACROBAT

This is the quickest, simplest and easiest method both to make corrections, and for your corrections to be transferred and checked.



1. Click **Text Edits**
2. Select the text to be annotated or place your cursor at the insertion point and start typing.
3. Click the **Text Edits** drop down arrow and select the required action.

You can also right click on selected text for a range of commenting options, or add sticky notes.

SAVING COMMENTS

In order to save your comments and notes, you need to save the file (**File, Save**) when you close the document.

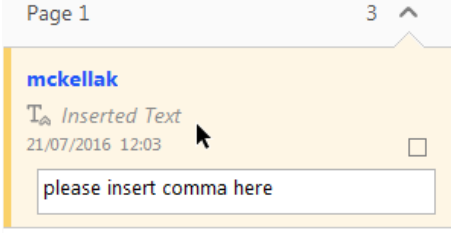
USING COMMENTING TOOLS IN ADOBE READER

All commenting tools are displayed in the toolbar. You cannot use text edits, however you can still use highlighter, sticky notes, and a variety of insert/replace text options.

POP-UP NOTES

In both Reader and Acrobat, when you insert or edit text a pop-up box will appear. In **Acrobat** it looks like this:

In **Reader** it looks like this, and will appear in the right-hand pane:



DO NOT MAKE ANY EDITS DIRECTLY INTO THE TEXT, USE COMMENTING TOOLS ONLY.

Oxford Journal of Legal Studies, (2018), pp. 1–30
doi:10.1093/ojls/gqy039

Social Norms and the Internal Point of View: An Elaboration of Hart's Genealogy of Law

Philip Pettit*

Abstract—HLA Hart tells us about how law would have emerged in a world of primary rules—informal, beneficial norms—by adjustments that the primary rules would naturally require; these adjustments would have introduced secondary rules for regulating the primary. But he does little to explain how the primary rules would themselves have emerged and, by most accounts, does not expand appropriately on the idea that the relevant players in his story would have taken an internal point of view, as he calls it, on the rules involved. This article, presented as the Hart Memorial Lecture on 14 June 2018, argues that both problems can be resolved at once. Elaborating on Hart's story with an account of how primary rules might have emerged, it also explains why those regularities would have attracted such a point of view. It argues that the emerging primary rules would have come to be perceived, internalised and ratified by all, and would even have been seen as rules to which each was personally committed.

Keywords: primary rules, secondary rules, genealogy, law, Hart, internalisation

1. Introduction

A. Hart's Methodology

One of the most remarkable things about Herbert Hart's classic study *The Concept of Law*, first published in 1961, is that without advertising the fact, perhaps even without recognising it explicitly, he introduces a way of doing

* LS Rockefeller University Professor of Human Values, Princeton University and Distinguished Professor of Philosophy, Australian National University. Email: ppettit@princeton.edu. I am most grateful for the enormously helpful comments I received at and after presenting versions of this article, first as the Hart Memorial Lecture in Oxford, June 2018, and later, as a colloquium paper, at Princeton University in September 2018. I also benefited greatly from comments on a draft of the paper provided by Emad Atiq, Nicola Lacey, Nelson Tebbe and the Journal's General Editor, Timothy Endicott.

philosophy that was quite at variance with the spirit of that time in Oxford.[1] It was a time that had been dominated in philosophy by JL Austin, who died in his late forties, just the year before Hart's book appeared. In the style of the period, Hart presented his work as an analysis of the concept of law. But he offered very little of the analysis of linguistic usage, common or professional, that we might have expected; in that regard, he broke from the ranks of his colleagues in Oxford and in the philosophical world in which Oxford then stood supreme.

What Hart offered instead might be described as a philosophical genealogy of the concept of law: an account of how it would plausibly, even inevitably, have gained currency in a society that previously lacked the concept—or, indeed, the practice—that it might have been used to articulate.[2] He begins with an imagined society that we might describe as Normitania, where informal norms of conduct have emerged among its residents: these he calls primary rules. He then shows how various difficulties would have prompted a set of adjustments among the people involved and how these, without any central contract or planning, would have given rise to secondary rules governing the interpretation, amendment and application of primary rules. He invites us at that point to agree that, by almost any standards of interpretative practice, it would be natural to cast this regime of primary and secondary rules as a system of law: that it deserves the name, to coin a term, of Lexitania.

The idea, as I read him, is that if the Lexitanians had a phrase for that combination of rules, as they surely would have done, it would correspond in our language, by natural criteria of interpretation, to the notion of a system of law. If we go along with that idea, then there are a number of benefits in prospect. We may take the story about how Normitania leads to Lexitania to direct us to the sort of thing that deserves to be conceptualised as a system of law and to how it can get to be conceptualised as such among its officials and subjects. We may take it to illuminate the role or function of the system, explaining why it is a more or less inevitable feature of a certain sort of society. And we may take it to demystify law in the positivist fashion that attracted Hart; the story does not have to appeal at any point to natural law or moral value.[3]

Hart's methodology harks back to the 18th-century idea of a conjectural history; it has parallels, for example, with the story that David Hume tells

[1] Famously, Hart casts what he is doing as an exercise in 'descriptive sociology': HLA Hart, *The Concept of Law* (3rd edn, OUP 2012) vi. For an argument that his project does not belong with ordinary sociology, or indeed with the analytical philosophy of his day, see L Green, 'Introduction' in Hart (ibid) xlv–xlvii. For a vivid account of his close relationships with his philosophical contemporaries, see N Lacey, *A Life of HLA Hart: The Nightmare and the Noble Dream* (OUP 2004). It is worth noting, of course, that some of his other work (for example, on responsibility) displays clear signs of Austin's influence.

[2] J Gardner, 'Why Law Might Emerge: Hart's Problematic Fable' in D d'Almeida, J Edwards and A Dolcetti (eds), *Reading HLA Hart's The Concept of Law* (Hart Publishing 2013).

[3] Note, however, that to allow for versions of positivism developed in the wake of Hart's work, it is not blocked from appealing to people's subjective judgments or valuations.

about the emergence and utility of the concept and custom of ownership.[4] It has come back into vogue in recent times, as when Edward Craig offers us a conjectural genealogy of the concept of knowledge,[5] and Bernard Williams a similar genealogy for the ideas of truth and truthfulness.[6] I have relied on it myself to provide a genealogy of ethical practices and concepts.[7] In this usage, the word refers to a counterfactual exercise, and should be distinguished from genealogy in a historical sense. Equally, it should be distinguished from the debunking sort of genealogy offered, for example, in Nietzsche's *Genealogy of Morals*; a debunking genealogy of any norms, legal or moral, would deprive them of the motivating and related power they may have seemed to possess.[8]

As I construe the genealogical method, it breaks with conceptual analysis in its standard form, but serves a similar purpose.[9] ~~Unlike~~, standard analysis, it does not itemise the connotations of the target term or concept—in this case, that of law; filter out those connotations that we would want to see preserved across any cases where we applied the term; and then, relying on this method of cases, seek to identify the property identified by the stripped-down set of connotations. Rather, it describes how an interpretatively equivalent concept might have emerged in tandem with a corresponding practice—how concept and practice might have co-evolved—and invites the conjecture that our own concept of law serves a similar role and identifies a similar referent: a system of primary and secondary rules.

B. The Aim and Plan

My aim in this article is not to defend Hart's genealogical method or narrative, only to extend it into the past, spelling out a story that his narrative presupposes. This is the story, cast in the same genealogical mould, of how a society might give rise to the primary rules that he postulates. In other words, it is a story about how Normitania might have emerged in a world without norms: in Prenormitania, as we may describe it.

The developed story extends the temporal reach of Hart's genealogy, taking it back to the earlier stage where primary rules are supposed to emerge spontaneously. In doing so, it evades a familiar problem raised for the spontaneous emergence of norms of a kind with the emergence that Hart presupposes. ~~Assume~~, as all sides do, that norms require enforcement. Why should people who are not altruistic enough to conform to norms of their own accord be altruistic enough to practice enforcement, bearing the associated

AQ1

[4] D Hume, *A Treatise of Human Nature* (OUP 1978) s 3.2.2.
[5] E Craig, Knowledge and the State of Nature (OUP 1990).
[6] B Williams, *Truth and Truthfulness* (Princeton UP 2002).
[7] P Pettit, The Birth of Ethics: Reconstructing the Role and Nature of Morality (OUP 2018).
[8] F Nietzsche, *On the Genealogy of Morals* (CUP 1997).
[9] Pettit (n 7); P Pettit, 'Analyzing Concepts and Allocating Referents' in A Burgess, H Cappelen and D Plunkett (eds), *Conceptual Engineering and Conceptual Ethics* (OUP 2019).

costs of effort or embarrassment, for example? The problem can be resolved if enforcement does not require costly intentional action, but only observing how someone behaves, and being seen by that person to observe them. The story told here, which is broadly faithful to Hart, shows how enforcement can take that non-intentional form and vindicates Hart's assumption about the emergence of primary rules in Prenormitania.

But a good part of the interest of providing a prequel to Hart's genealogy is that doing so supports a rich psychological model of what it means for people to take an internal point of view, as he calls it, on the rules that they follow. According to Hart, the internal point of view is that which is held by people who are insiders to the society, not outside observers, and in particular insiders who willingly abide by the rules: they do not submit only in the manner of detached outliers.[10] The genealogy of primary rules developed here identifies different dimensions in an internal point of view: different ways in which it may contrast with the view that outside or detached observers will take of rules. It extends the psychological depth of the genealogy, as it extends its temporal reach.

While the story focuses only on primary rules, ~~of course,~~ the account it offers of the internal viewpoint that people in Prenormitania will adopt on those rules naturally extends, with suitable amendments, to the secondary rules on which Hart concentrates.[11] According to the story, primary rules will not only get to be established as universal regularities of behaviour in Prenormitania; in terms to be explained later, they will also come to be matters of perception, internalisation and ratification, and will be registered in the process as matters of personal commitment. Norms that achieve that status will count in an intuitively rich sense as enjoying an internal status in the perception and psychology of agents.

While the discussion that follows is meant to be faithful to the spirit of Hart's treatment of law, providing an extension that I hope he would find congenial, it is not designed as an exegesis or defence of his views. Thus, it ignores the vast literature that explores and often questions those views. It focuses only on the two aims cited: first, to tell a story about how the primary rules that he postulates might have emerged spontaneously in Prenormitania, as he thinks secondary rules would have emerged in Normitania; and, secondly, to show

[10] Hart (n 1) 89–92.

[11] Hart (n 1) 90. Hart assumes that in the case of secondary rules the internal point of view will involve a 'much extended and diversified' vocabulary, with new terms appearing like 'jurisdiction' and 'validity'. He suggests in that passage that primary rules will attract the internal point of view for most people in a 'simple regime of primary rules'. But he later allows that secondary rules may have to be seen from the internal point of view only by officials (115–17). Indeed, he seems to take this to be enough also for various first-order rules, on a par with primary rules, that are endorsed under secondary procedures; that may be because such rules need not be collectively beneficial in the manner of spontaneously supported primary rules: they may be introduced under minority pressure. None of these features undermines the relevance of the extension to his genealogy that is proposed here.

how essential to that process would have been the adoption of an intuitively internal point of view on the emergent rules.

In the conclusion at the end of the article, I look briefly at the main remarks that Hart makes about primary rules and argue that they apply to the sorts of social regularities or norms explored here. But the point in this exercise is to vindicate the claim that the role he ascribes to primary rules in grounding a legal system is precisely the sort of role played by the social norms introduced here. It is not to suggest that the genealogy provided for those norms is guided by reflection on Hart's comments.

Apart from that conclusion, the argument of the article is developed in three main sections. In the first, I draw on familiar stories about the emergence of social regularities to explain why we would expect universal, reputationally supported patterns of cooperative conduct—collectively beneficial norms—to appear in Prenormitania; other, objectionable norms are liable to have appeared there as well, of course, but they are not relevant to this project. In the second, I argue that, in all likelihood, those cooperative norms ought to achieve a distinctive social status among the population: as a matter of common awareness, they ought to be perceived as norms by members; to be internalised by them, not treated as making unnecessary and unwelcome demands; and to be ratified communally, as the members present them to youngsters and newcomers as norms that they will be expected to follow. The third and final section takes this line of thought further; it argues that there is a perfectly ordinary, pre-moral sense of commitment in which the members of the society will be personally committed to conforming to those cooperative norms, granting others a claim to censure any failure on their part to do so.

2. *The Emergence of Universal Patterns of Cooperative Behaviour*

A. Some Assumptions

There are a number of assumptions that we have to make about Prenormitania in order to explain the emergence of primary rules there. Many of these are so obvious that they may pass without remark, but five are particularly salient. They are all fairly plausible, and Hart himself endorses them implicitly, ~~and~~ sometimes explicitly, in his own story about the transition from Normitania to Lexitania.

A first assumption is that the members of this society are self-regarding to the extent that they are especially, if not exclusively, concerned about their own welfare and the welfare of those close to them. A second is that they are relatively rational in adopting measures for the promotion of that welfare: reliable in forming beliefs about available opportunities, for example, and in selecting appropriate means of action. And a third, more specific assumption is

that they have the use of natural language. However rudimentary this language is, it allows them to communicate in the standard, distinctively human manner.[12]

While these first three assumptions are only implicit in Hart's discussion, he explicitly mentions another two. One is that the members of the society he explores enjoy 'approximate equality' with one another. None is 'so much more powerful than others, that he is able, without cooperation, to dominate or subdue them for more than a short period'.[13]

This is the most problematic of his assumptions, since most societies in history have been hierarchical, with one class or ethnicity or gender effectively dominating others; they constitute the alphas in a world where there are also betas and gammas.[14] There are two possible defences for his approach, however, one quasi-philosophical, the other quasi-historical. The more obvious, quasi-philosophical defence is that, since Hart is looking at how the rules of a presumptively inclusive society might have emerged—and this necessarily in a counterfactual way—he can idealise the starting point without jeopardising the methodology described earlier.

The quasi-historical defence may be called upon if such idealisation is thought inappropriate. It consists in the claim that all that the genealogy needs strictly to suppose is that there is one dominant class in the society, the alphas, among whom approximate equality obtains; that members of this class are likely to support cooperative rules, both primary and secondary, governing their interactions with one another; and that there is nothing to stop those rules from extending to other classes in the society, if members of those classes manage to achieve more or less equal power with the alphas: this might come about spontaneously or, more likely, as a result of resistance and struggle. The cooperative rules that the genealogy would hold ~~out then as a model in the interpretation of a law are~~ allowed to emerge, as they undoubtedly emerged in the actual world, as rules governing primarily a privileged, relatively equal class: this, at any rate, provided there is no obstacle to their being extended to become rules governing everyone equally.[15]

[12] TC Scott-Phillips, *Speaking Our Minds: Why Human Communication is Different, and How Language Evolved to Make it Special* (Palgrave Macmillan 2015). Speakers overtly convey their would-be attitudes to one another—how they take things to be, what they desire or intend to do—by advertising that that is indeed what they wish to convey. On this, see P Grice, 'Meaning' (1957) 66 Philosophical Review 377; P Grice, *Studies in the Ways of Words* (CUP 1989); S Neale, 'Paul Grice and the Philosophy of Language' (1992) 15 Linguistics and Philosophy 509. They will do this truthfully if they actually have those attitudes, deceptively if they do not.

[13] Hart (n 1) 195.

[14] This is certainly true of societies since the time of the agricultural revolution about 10,000 years ago; on the possibility of more equality in earlier groups, see C Boehm, *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Harvard UP 1999); C Boix and F Rosenbluth, 'Bones of Contention: The Political Economy of Height Inequality' (2014) 108 American Political Science Review 1. Hart (n 1) relaxes the assumption of equality in his discussion of international law.

[15] In any hierarchical society of the kind envisaged in the genealogy, there would certainly be subjugating norms in place, expressing and enforcing the dominance of the alphas. But the idea is that such norms would be undone, and cooperative norms would be extended, if other people managed to gain equality with the alphas. To assume that there is a class of alphas among whom approximate equality holds is itself to idealise somewhat,

Whether defended in the more philosophical or more historical manner, we will go along with Hart's assumption that Prenormitania is a society of equals in the attempt to elaborate his genealogy. Interpreted on the lines suggested, the assumption need not be as historically insensitive as it might otherwise seem.

The fifth and final assumption that his and, indeed, our story needs is also more or less explicitly endorsed by Hart. This postulates that the members of the society are each dependent for their well-being on others reliably eschewing violence and deception and the like: on their practising what he casts as 'mutual forbearance and compromise'.[16] The assumption is that in Prenormitania it will be manifest to each member, as it will be manifest in any human society, that each needs to be able to rely on others for such basic forms of tolerance and cooperation.

B. Reliance, Reputation and Regularities

With these assumptions in place, it becomes intelligible why, for purely self-interested reasons, each member of the society should display certain patterns of cooperative behaviour in dealing with others. Human beings may well be spontaneously cooperative, given the selectional advantage of such a disposition.[17] But while that disposition would increase the likelihood of cooperative behaviour in Prenormitania, we need not rely on its presence for postulating cooperation; the emergence of cooperative behaviour will be intelligible even if it is absent. Nor, for similar reasons, need we postulate that Prenormitanians are spontaneously disposed, for selectional reasons, to punish any defector by refusing to cooperate with them;[18] we need not rely on this possibility, even if it would help the case to be made for cooperation.

To start with a salient example, Prenormitanians may be expected to be truthful or sincere in making reports to one another and to be relatively careful about the reports they make. Each will depend on receiving reliable reports from others about features of the environment with which they are not immediately familiar: say, about whether the seasonal fish are returning to the river, whether the hunting prospects are better up north, whether the wild corn has begun to ripen, and so on. But if I am not reliable in the information I

holding out the possibility of a rule of cooperative norms before which all can aspire to be equal; it represents a more optimistic scenario than one where a single family dominates. The account later provided of how primary rules or social norms are likely to develop appeals to the economy of esteem, and this economy can explain the appearance of such subjugating norms: as honour among thieves may require hurting others, for example, so the same may be true of honour among alphas. See G Brennan and P Pettit, *The Economy of Esteem: An Essay on Civil and Political Society* (OUP 2004).

[16] Hart (n 1) 195.

[17] M Tomasello, *Why We Cooperate* (MIT Press 2009); M Tomasello, *A Natural History of Human Morality* (Harvard UP 2016).

[18] E Fehr and S Gächter, 'Altruistic Punishment in Humans' (2002) 415 Nature 137.

provide to you about such matters, I can hardly hope to find you reliable in providing information for me. And equally, I can hardly hope to get you to rely on me in any future scenario—to believe what I say—where it may be in my interest to secure your reliance.

Why should you bother to prove reliable when I seek information from you, since my proven lack of reliability will deprive you of a self-interested incentive to report reliably to me? Indeed, my lack of reliability may even provide you with a positive incentive to penalise me by not proving reliable; you may plausibly expect this to get me to rethink my attitude: to recognise that in exchanges with you, honesty is the best policy. And why, even more tellingly, can I expect to be able to get you to rely on me in the future, accepting what I say, if I have proved unreliable in the past? Once bitten, as the proverb goes, twice shy.

What holds in these regards for me, of course, will hold by the same token for you. And by that same token, it will hold for anyone who is likely to relate to another in a similar way: to relate, now as someone in need of information, now as someone in a position to provide it; now as someone anxious to get another to rely on them, now as someone whose reliance the other desires. In effect, it will hold for everyone insofar as everyone is bound to operate within relationships of mutual dependence with others.

Summing up this lesson, then, we Prenormitanians are each bound to be concerned to have a reputation among those with whom we interact for reliability as reporters. And this concern is bound to prompt us, at least in the general run of cases, to report truthfully and carefully to one another. In that run of cases, there will be little or no difficulty in telling the truth; there will not be particularly huge benefits, personal or even collective, to be won by deception; it will not be easy to hide our deception from those we mislead; and we will not be subject to impulses or passions that would overwhelm our recognition of the case for proving reliable.

Reliable truth telling has two salient characteristics: it is a collectively beneficial pattern or regularity in the sense that everyone is better off in any society to the extent that everyone reliably tells the truth to others; and yet it is an individually burdensome pattern in the sense that each may be better off still if they can rely on others to tell them the truth, while doing so themselves only when it suits their interests. The observations about truth telling highlight the likelihood that, despite its being individually burdensome in some cases, Prenormitanians will be generally disposed to support such a collectively beneficial pattern.

The points made in support of this conclusion apply also with other collectively beneficial, individually burdensome patterns or regularities. They argue that as you and I, and other members of the society, will generally tell the truth, so will we be generally non-violent in our behaviour towards one another, non-coercive in seeking to elicit one another's responses, faithful to

the promises we make and, if there is a sense of ownership in place, respectful of property. Despite being collectively beneficial, any such pattern will be individually burdensome: we are each likely to wish that we could get away with breaching it opportunistically, while others reliably conform. Still, the reputational points made in support of the claim that Prenormitania is a truth-telling society argue equally that it will be a society where non-violence, non-coercion, promise-keeping, and so on are common patterns of behaviour. That they are behavioural patterns of this kind does not yet mean, of course, that they count as norms, let alone social norms: more on this in a moment.

C. Multilateral, Reputational Support

The reputational mechanism on which this story of cooperation relies generates a pattern of bilateral interaction and cooperation that resembles the pattern described in game-theory as tit-for-tat reciprocation.[19] This pattern applies whether the members in a pair of interactants act simultaneously in each encounter, take turns in acting or sometimes act simultaneously, sometimes take turns. A person follows the tit-for-tat strategy just in case they cooperate in the first interaction—they tell the truth, avoid violence, coercion and theft or keep their promises—and in every later episode they do whatever the other party did in the previous encounter with them: they cooperate if the other cooperated, they defect if the other defected.

Strictly speaking, tit-for-tat patterns might emerge in suitable forms of behaviour—say, in avoiding violence or even in providing assistance—in the absence of language; the reputational mechanism required does not require communication. It is not surprising, then, that there is some evidence that such patterns materialise among the members of many different species, not just amongst human beings.[20] But the presence of language means that the reputational mechanism can assume a particularly powerful, multilateral form. It can generate an encompassing economy of esteem.[21]

The desire to be thought cooperatively reliable by a partner gives you an incentive to cooperate, in the bilateral model considered up to now, insofar as such behaviour manifestly provides the other with evidence of your disposition to be cooperative with others—or at least to be cooperative with those who cooperate with you. But the presence of language—or, more generally, of a means of communication—has an effect that transforms the model, giving it a multilateral rather than just a bilateral character.

[19] It only resembles tit-for-tat, since it need not make requirements of the exacting kind imposed in game-theory models. It need not require retaliation in response to any single defection, for example, and need not require a continued pattern of retaliatory defection until the other begins to cooperate. R Axelrod, *The Evolution of Cooperation* (Basic Books 1984).

[20] M Nowak, 'Five Rules for the Evolution of Cooperation' (2006) 314 Science 1560.

[21] Brennan and Pettit (n 15).

Language will enable you to give evidence of wanting to be reliably cooperative yourself by means other than actual cooperation. Giving favourable or unfavourable reports to third parties on the performance of others whom you can identify by name or description will provide evidence of that kind. If you gossip about the failure of others, then you manifestly stick your neck out, as we say: you invite the charge of hypocrisy, not just uncooperativeness, if you fail to prove reliable yourself. And by putting more at reputational stake in this way—by linking failure to a greater cost—you communicate all the more credibly your claim to be a cooperative, reliable agent yourself. By gossiping about the failures of others to third parties, you will invite them to think that, running the risk of being named a hypocrite if you fail to cooperate, you must definitely be a reliable cooperator.

In view of the impact of language, Prenormitania will be a gossipy society, in which you and I realise that if we let down someone in an interaction, that may have multilateral as well as bilateral costs. The betrayal may not only lead the person you let down not to cooperate with you in future; it may also deny you the possibility of securing the cooperation of others among whom word of your defection has spread. Language will allow your victim's bad opinion of you to propagate and gain wider currency, giving you a bad name: a standing reputation for being unreliable.

This reputation may not only come to be shared widely in your local networks, or in the society at large, it may even come to be shared widely as a matter of common awareness. This will happen if the evidence leads each, not just to form a bad opinion of you, but to become aware that each holds such a bad opinion, to become aware that each is aware in turn of this, and so on.[22] At that point, you will not only be esteemed badly by each, suffering a consequent degree of shame; you will be badly esteemed in the public manner associated with stigma.

D. Regularities and Norms

The multilateral reinforcement of the reputational mechanism will give rise to an economy of esteem insofar as each of us in Prenormitania gains a particularly powerful motive for seeking to behave cooperatively in the general run of cases. These are cases, as we characterised them earlier, where there are no particular difficulties that block our cooperative behaviour, there are no enormous benefits to be won by defection, it is not easy to hide defection and we are not overwhelmed by countervailing impulses or passions. In any such case, each of us may be tempted to defect on the particular person we are dealing with: the bilateral reputational incentive may not be powerful enough to ensure cooperation. But that temptation will be much weaker in the

[22] D Lewis, *Convention* (Harvard UP 1969). A more detailed account of common awareness is offered in the next section.

presence of the multilateral reputational incentive that comes onstream with communication and language. The concern for how we are likely to be seen by others in general—the concern with our good name—should serve to keep most of us onside in most relevant cases.

These observations support a claim that various collectively beneficial, if individually burdensome, patterns of cooperation will emerge in Prenormitania. Each such pattern or regularity, R, will satisfy broadly the following conditions, whether R involves reliably telling the truth, reliably avoiding violence or coercion, promise-breaking or theft.

Almost everyone in the community conforms to R, at least when dealing with those who have not breached R in a previous encounter.

Almost everyone expects such conformity to attract a favourable reputation among others and/or nonconformity to attract an unfavourable reputation.

Almost everyone is supported in their conformity to R by expecting such reputational benefits and costs.

The sorts of regularity that R exemplifies contrast with mere conventions, such as the convention of driving on the right or driving on the left hand side of the road.[23] Conventions, unlike the regularity of truth telling, are grounded in the attraction for each of us, including each of us in Prenormitania, of doing the same thing as others—securing a desired form of coordination—and need not attract or depend in any way on the expectation of reputational benefits. Thus, conforming to a convention is not individually burdensome, and does not give rise to freeriding problems, in the manner in which the regularities we are envisaging may do. You will not be tempted, on approaching another car, to break convention, drive in the same lane and risk a head-on collision.[24]

AQ25

The three conditions they satisfy mean that the sorts of regularity envisaged here are distinct not just from mere conventions, but also from a number of other social regularities that may arise in any society, including Prenormitania.

Thus, the fact that they are generally maintained in people's behaviour, as the first condition stipulates, distinguishes them from standards honoured more in the breach than in the observance, such as a supererogatory ideal that we routinely applaud but rarely realise. The fact that they reflect the mutually expected attitudes of agents toward conformity, as the second condition holds, means that they are distinct from regularities to which others are manifestly indifferent, such as the regularity whereby most people sleep at night, not during the day. And the fact that they are supported by that expectation, as the third condition holds, means that they are distinct from regularities such as that involved in taking steps to guard against penury in old age; it is unlikely

[23] ibid.

[24] E Ullmann-Margalit, *The Emergence of Norms* (OUP 1977). Leslie Green ~~(in Hart~~ (n 1) xxii–xxiii~~)~~ argues rightly that Hart is mistaken to suggest in a later comment that his theory applies mainly to rules that have a conventional, coordinating aspect. Individually burdensome rules, on which *The Concept of Law* focuses, are not mere devices of coordination, like the convention governing the side of the road on which to drive.

that people are motivated to display such prudence by expecting that others will think well of them for doing so.

How to characterise the universal patterns or regularities captured in our conditions? For want of a better term, we may describe them as norms that obtain among the members of the society. But as we shall see in the ~~second~~ section, there is good reason to regard them as pre-social rather than properly social norms. For all that has been said here, as we shall see, they may not be generally perceived as norms to which almost everyone conforms; the regularities involved may obtain without this being registered by the people who sustain them.[25]

AQ3

E. The Role of Excuses

Before moving on to the discussion of perception, it may be useful to mention one feature of the pre-social norms that we might expect to emerge in Prenormitania. This is that it is bound to be possible within an economy of esteem for someone to breach a norm, be seen to breach the norm and yet avoid the reputational costs of doing so.

Suppose, for example, that I mislead you by a report that the blackberries on the other side of a local hill have ripened. It may be that you go in search of them and find they are still not fit for picking or find instead that they have all been picked already. And imagine that I convince you either that they really looked ripe to me—this, perhaps, because I had seen them in the setting sun—or that someone picked them between the time I observed them or made my report and the time of your search. Imagine, in short, that I can credibly explain my having misled you either by appeal to ~~a~~ misleading evidence or a changed world.

In such a case, it would be self-defeating of you to treat me as if I had breached the norm of truth telling by not being careful or truthful in the report; it would lead you to expel me unnecessarily from the community of those you can generally rely on. And so, we may expect you to keep me within that circle if you accept a misleading-evidence or changed-world explanation.

[25] Anticipating discussion in the next section, social norms must plausibly satisfy the three clauses given, plus a fourth clause to the effect that the first clause, or even all three clauses, are fulfilled as a matter of common awareness. For an earlier version of this conception of social norms, see P Pettit, 'Virtus Normativa: Rational Choice Perspectives' (1990) 100 Ethics 725, reprinted in P Pettit, *Rules, Reasons, and Norms* (OUP 2002) and Brennan and Pettit (n 15); the current version appears in P Pettit, 'Value-mistaken and Virtue-mistaken Norms: Political Legitimization without Morality?' in J Kuehnelt (ed), *Political Legitimization without Morality?* (Springer 2008). See also P Pettit, *The Robust Demands of the Good: Ethics with Attachment, Virtue and Respect* (OUP 2015); Pettit (n 7). The notion of a social norm picks up points made in a variety of approaches. See eg P Winch, *The Idea of a Social Science and Its Relation to Philosophy* (Routledge 1963); J Coleman, *Foundations of Social Theory* (Harvard UP 1990); E Sober and DS Wilson, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Harvard UP 1998); J Elster, *Alchemies of the Mind: Rationality and the Emotions* (CUP 1999); S Shapiro, *Legality* (Harvard UP 2011). For an insightful development of the idea of reputationally supported norms, see KA Appiah, *The Honor Code: How Moral Revolutions Happen* (Norton 2010). For an overarching theory that is reconcilable with that adopted here, although it uses terminology somewhat differently, see G Brennan and others, *Explaining Norms* (OUP 2013).

In virtue of having this effect, the explanation counts as an excuse; it enables me to get off the reputational hook.

The observation suggests that while the economy of esteem will motivate people in Prenormitania to act in a way that establishes a norm of truth telling—and, of course, norms of non-violence and the like as well—it will operate in a relatively nuanced way that leaves room for letting those who breach such norms occasionally off the hook. This observation about excuses will be important later, in our discussion of ratification. It applies straightforwardly to full excuses of the kind illustrated, but also extends to partial excuses, as we generally call them; to excuses of a distinct non-epistemic kind such as 'I was coerced to say what I did'; and, indeed, to the exemptions relevant in cases where someone is not functioning properly as a person.[26]

3. The Perception, Internalisation and Ratification of Cooperative Norms

For all that we have said about the norms that are likely to emerge and stabilise in Prenormitania, it should be clear that they are not likely to be seen by agents from within anything like Hart's internal point of view. That point of view presupposes that those who conform to the rules see them as general rules, and take an attitude towards them that contrasts with that of an outsider or outlier. But, for all that has been said up to now, Prenormitanians may not be able to see the norms to which they conform as general rules of this kind and may not be able to adopt any attitude whatsoever towards them.

A. Perception

The cooperative norms whose emergence is documented in the previous section may constitute universal patterns that escape the notice of those who sustain them. Adam Smith famously argued that various aggregate economic regularities may emerge behind people's backs 'by an invisible hand'; strictly, he should have said by a not necessarily visible hand.[27] The story told about the emergence of norms of truth telling, non-violence, non-coercion and the like has a similar character. For all it stipulates, the patterns that emerge as a result of people's each seeking to have the reputation for being reliable cooperators may be regularities of the same, unnoticed kind.

You and I, and everybody else in Prenormitania, may be led by reputational pressures to cooperate reliably in the domains illustrated without anyone

[26] See J Gardner, *Offences and Defences: Selected Essays in the Philosophy of Criminal Law* (OUP 2007). The observation connects with some of the comments made by Richard Holton on how legal rules may have exceptions, although he concentrates on cases where a failure to stick by a rule is not so much excusable as justifiable: this, in virtue of intuitively more important considerations trumping the case for abiding by the rule. See R Holton, 'The Exception Proves the Rule' (2010) 18 Journal of Political Philosophy 369.

[27] A Smith, *The Theory of the Moral Sentiments* (Liberty Classics 1982) 184–5.

noticing that this is what more or less everyone in the society does: that it constitutes a more or less universal pattern. Confronted now with one interlocutor or interactant, now with another, you may confidently expect that person to be cooperative; in that sense, you may believe of everyone you confront, or are even liable to confront, that they are cooperative. But you may not register this as a general truth; you may not register the truth of the proposition that everyone, or almost everyone, is cooperatively disposed.

For all that has been said so far, then, it may be true that, for any rule or norm like that of truth telling or non-violence or whatever, you as an agent in the society will prefer in any instance to act in a way that happens to involve conforming to the rule; you will have a powerful reputational incentive to act in that manner. But it need not be true that you prefer to conform to the rule or norm under that description of what you do: you may not be aware that there is a general rule or norm to conform to.

It is highly likely, however, that as a regularity like truth telling or non-violence or whatever gets established in a society, it will come to be perceived as an aggregate fact about how people behave in the society: the evidence of its obtaining will be there for all to see. Indeed, it is likely that it will come to be perceived in this way as a matter of common awareness among the members of the society. Not only will the evidence be there for all to see; it will be there for all to see that the evidence is there for all to see; it will be there for all to see that it is there for all to see that the evidence is there for all to see; and so on. In short, the evidence will not just be accessible, but accessibly accessible, accessibly accessibly accessible, and so on in a recurring pattern. The universal regularity of truth telling or non-violence or whatever will be established as a matter of common awareness: each will be aware that it holds, aware that each is aware that it holds, aware that each is aware that each is aware of this, and so on in the usual hierarchy.[28]

That a norm in the sense characterised earlier is perceived by all, as a matter of common awareness, means that it deserves to be cast as a properly social norm. It will be clear to you, as it will be clear to each, that conformity is the local norm, as we say. It is the norm in the statistical sense of being a routine, expected pattern in the society. And it is the norm in the regulative sense of constituting a *sine qua non* condition for anyone hoping to gain acceptance in the society: fail to live up to it, and you will count as an outsider or an offender.[29]

[28] Lewis (n 22). This does not require that everyone is actively aware at each level in the hierarchy of the claim that common awareness supports there; that would be impossible for any finite mind. All that it need require is that each is disposed, should they confront and understand the relevant question at any level, to answer in the affirmative.

[29] The argument for its being a matter of common awareness that a regularity exists suggests that it will also be a matter of common awareness that the other two clauses in our characterisation of a norm are satisfied. Similar observations about the evidence accessible to each party, accessibly accessible to each, and so on, apply with these clauses as well as with the first. Thus, in all likelihood, it will not only be the case that it is a matter of common awareness that almost everybody conforms to the regularity, R; it will also be a matter of common

At this point in the genealogy, it may be thought that the internal point of view makes a first appearance. You will not just have a reputational incentive to act in a way that happens, perhaps unbeknownst to you, to involve conforming to a norm like that of truth telling; you will now have a reputational incentive to conform to the norm as such: to conform consciously or knowingly to the norm. Thus, it will at least be possible for you and others to have a general attitude towards the norm, and the question is whether acting on the reputational incentive means that that attitude will constitute an internal point of view.

The reputational incentive on which you act will be that if you do not conform with the universal pattern involved, then you will not be one of the gang; your social identity and allegiance will be in doubt. You will not be in tune with how things, as we others may say, are done around here. This suggests that, recognising the incentive on offer, you will not view the norm from an outsider or outlier perspective; you will display the sort of profile that Hart associates with the internal point of view.

But does the suggestion stand up to reflection? Why think that having a self-interested, reputational incentive for conforming knowingly to the norm marks off your attitude from that of an outlier who conforms unwillingly to it? After all, the unwilling outlier in a community might also conform knowingly to a norm and might have a self-interested incentive for conforming.

Nothing much hangs on how we decide this question. But a possible reason for thinking that the internal point of view already enters at this point is that there is a salient difference between acting to avoid a reputational penalty for being out of line with others and acting to avoid, say, a physical penalty for being out of line. A concern with the reputational penalty involves a concern for how you stand in the opinion of others, whereas a concern with the physical penalty may not involve anything of that character.

Even if the self-interested, reputational concern that leads you to abide by a norm like that of truth telling or non-violence is distinct in this way from the fear of a physical penalty—and even if it is taken to mark the first appearance of an internal point of view—it has one feature in common with the fear of a penalty that is worth noting. This is that it has a causal-instrumental character. If you fail to conform in this or any instance, then the consequence you fear is a downstream, causal effect: viz, that you will get a name for being unreliable in relevant exchanges and, more generally, for not accommodating recognised expectations.

While the causal-instrumental consideration is more or less bound to support conformity in this way, however, it is important to notice that the consideration

awareness that almost everybody will expect conformity to attract a positive reputation and/or non-conformity a negative, and that this expectation will support people's tendency to conform.

may be relevant without being the conscious, motivating reason why you or anyone else generally conforms. You and others may be actively moved, not by self-interested calculation, but by sheer habit or even by more idealistic considerations. And yet it may remain the case that the causal-instrumental consideration plays an important role in supporting your compliance.

This will happen in an obvious way if the self-interest is effectively motivating only with the motivational help of habit or idealism. But it may also happen, and may be more likely to happen, even when habit or idealism are the only motives driving the behaviour. In this case, the causal-instrumental consideration can play an important back-up role, being there to prompt you to conform in the event that the more regular motive fails: in the event, for example, that habit falters or idealism wanes. The consideration may be salient enough to alert you or any would-be defector to the self-interested costs of defection and to put you back on track. It may monitor your behaviour, coming into play if, and only if, it is needed; it may make it robust over scenarios where the usual generators of conformity fail.

B. Internalisation

We have seen that a norm may obtain in a society without being perceived as a norm and that it is only with perception that it deserves to be cast as a properly social norm. But the next, surprising point to notice is that a norm may be perceived as a norm, and be in that sense established, without being internalised among people: this, on a more or less intuitive understanding of what that involves.[30] Internalisation connects with Hart's notion of the internal point of view, but it is not exhaustive of the idea, as will be shown by the account presented later.

In order to see what internalisation means, and why it need not materialise with some norms, it may be useful to consider the phenomenon of pluralistic ignorance, as it has been described by two psychologists, DA Prentice and DT Miller.[31] In an experiment that they conducted on a university campus, they found that while the students they studied generally conformed to a regularity of drinking a certain amount of alcohol at parties and other events, while they generally expected that conforming to that regularity was essential for maintaining a suitable reputation in the eyes of others and while they were therefore generally motivated to drink the prescribed amount, the expectation that each of the others would disapprove of a failure to conform was mistaken. In fact, most of the students thought that the amount in question was excessive

[30] RD Cooter, 'Structural Adjudication and the New Law Merchant: A Model of Decentralized Law' (1994) 14 International Review of Law and Economics 215.
[31] DA Prentice and DT Miller, 'Pluralistic Ignorance and Alcohol Use on Campus' (1993) 64 Journal of Personality and Social Psychology 243.

and would have preferred not to be constrained by the norm they combined to sustain.

The drinking norm in operation among the students in this example is a genuine norm, by our account. It is a regularity that meets all the conditions in our earlier characterisation, since almost everybody conforms, almost everybody expects conformity to be essential for reputation and almost everybody is motivated by that expectation to conform. Moreover, plausibly, it is a regularity that obtains as a matter of common awareness in the student body involved: it is a perceived norm, established as a socially routine and socially regulative pattern.

But however far it qualifies as a norm, and however fully it is perceived as a norm on all sides, the drinking rule does not enjoy internalisation in the sense invoked here. It is true, as with any perceived norm, that each agent in the group has a causal-instrumental reason to conform, expecting to suffer a serious reputational cost in the event of not doing so; that expectation is supported by the perception that conformity is the socially routine, regulative pattern. But the striking thing is that while everyone recognises that conformity is the general pattern, and forms the corresponding expectation about the reputational stakes, no one apparently prefers that that pattern should obtain. Each would prefer, presumably, that the norm in place amongst them should require the consumption of a lesser amount of alcohol or perhaps allow the consumption of none at all.

This observation allows us to define internalisation. Any perceived norm that leads most or all members of a group to do X is going to be supported by the expectation among members that defection will have reputational costs, and compliance reputational benefits. But, as we can now see, members may not generally prefer that X-ing be the established pattern—they may each prefer a feasible, salient alternative, Y-ing—and, as a consequence, may not think ill of someone who fails to do X—or well of someone who does X—and thereby impose a reputational sanction. In that case, we can say, the norm will not be internalised. In order for the norm to be internalised, the members of the relevant group must generally prefer that X-ing be the established pattern—rather than any feasible, salient alternative—and must, in consequence, be disposed to think ill of someone who fails to abide by it and well of someone who lives up to it.

It may be that many norms obtain despite the fact of not being internalised in this sense by those who sustain them.[32] And it may be that some norms are internalised but only very reluctantly on the part of some: they may prefer that a norm obtains only because more appealing alternatives are infeasible. Norms of a kind that benefit some at a cost to others are unlikely to be internalised by those others, or unlikely to be internalised enthusiastically. But the cooperative

[32] Pettit, 'Value-mistaken and Virtue-mistaken Norms' (n 25).

sorts of norms at the focus of discussion here are not unbalanced in that manner, and are likely both to be internalised and internalised with enthusiasm.

Everyone in Prenormitania stands to benefit from people's generally conforming to norms like those of truthfulness, non-violence and non-coercion, keeping promises and, at a stretch, respecting property; this last case will be problematic to the extent that the local understanding of property is contested. People may each wish they could themselves breach those norms while others conform, but that alternative is not only infeasible; being a pattern that no one else could be expected to support, it hardly counts as an alternative at all. Among the alternatives in any area, then, they are going to prefer that everyone should conform to the sort of norm mentioned: truth telling, non-violence, non-coercion, fidelity to promises or respect for property.

The upshot is that the pre-social norms characterised in the first section are not only likely to be perceived as social demands in Prenormitania, but also to be internalised spontaneously by those who live under them; henceforth I shall take the internalisation they attract to be of the enthusiastic variety. Indeed, those norms are likely to be internalised as a matter of common awareness. This is because similar points to those that were invoked in earlier cases apply here. The evidence of internalisation—the clear appeal of the norms—will itself be evidently available to all. The fact that it is evidently available to all will itself be evidently available to all. And so on in the usual pattern.

The perception of a norm as a norm may already introduce an internal point of view among those to whom it applies, as tentatively suggested earlier, with each having a causal-instrumental, reputational reason for complying with the norm. But the internalisation of the norm certainly marks the advent of an internal point of view. It puts into play a new respect in which the viewpoint of members contrasts with the detached perspective of outsiders or outliers.

Insofar as you internalise a norm like truth telling or non-violence, you will not only have a causal-instrumental reason for conforming; you will also be responsive to the consideration that by conforming you contribute a necessary element to the universal pattern of truth telling or non-violence, where that is a pattern that you prefer to obtain, given that you cannot feasibly hope to freeride. Your conformity is necessary to that pattern, because if you did not conform, the pattern would not obtain. As a matter of its nature as an act of conformity, then, the action will have a property that makes it attractive to you: the property of making universal conformity possible. By conforming, you help to realise that property, where realising it is an immediate, constitutive effect rather than an effect of a downstream, causal kind.[33]

Because the realisation of that property is an effect of what you do—it is something you bring about by conforming—it still counts as an instrumental

[33] P Pettit, 'Three Mistakes about Doing Good (and Bad)' (2018) 35 Journal of Applied Philosophy 1.

reason for conforming. But it is a constitutive, not a causal, reason. And it is a reason that may not reflect any concern for your own interests; this is consistent with the genealogy, although not required by it. You may be inclined to act in the relatively disinterested way required spontaneously, rather than for any ulterior motive of self-interest. Or you may embrace that behaviour as a personal ideal: recognising that you would want to act on the disinterested inclination even if you were impeded by laziness, lassitude or distraction, you may be moved by that very consideration to overcome any such impediment.[34]

C. Ratification

As they have been characterised so far in the genealogy, cooperative norms of truth telling, non-violence and the like are bound to attract the support of members on the basis of self-interest, although they may also happen to appeal to their spontaneous inclinations or ideals, or answer to their unthinking habits; in that case, the self-interest will provide back-up reinforcement: it will monitor the behaviour of any member, being ready to play an active part in motivating compliance if, but only if, the independent inclination or ideal fails.

For all that the genealogy requires, then, the norms established will recommend themselves to people in a purely personal perspective; they are rules of behaviour that people each have their own personal reasons—albeit reasons of a similar kind—for adopting. Social ratification, as envisaged here, would introduce yet a further dimension to their way of viewing the norms. It would lead them to see the norms as standards that they support as a community and are obliged, for that reason, to honour as members.

The ratification of a norm presupposes its internalisation, as its internalisation presupposes its perception. To ratify a norm, by the account adopted here, is to accept that you cannot object to anyone's avowing a favourable attitude towards it in your name as a member of the group. In order to understand what ratification involves, then, it is essential to move away a little from the main line of argument and introduce the notion of avowal.

Avowal is a way of communicating an attitude that contrasts saliently with giving a report on the presence of the attitude. If I report on an attitude of my own, and I turn out not to display the attitude, then, as noted at the end of section 1 I can seek to excuse myself on the grounds that I was misled about my mind or that I changed my mind since making the report. Asked about whether I trust Jones, for example, I may report that I think that I trust Jones but am not quite sure. In that case, if it turns out that I later behave as if Jones is not trustworthy, I may excuse my miscommunication by saying that I must

[34] M Smith, *The Moral Problem* (Blackwell 1994). Instrumental reasons are often wrongly taken to be necessarily causal-instrumental and necessarily self-interested. The sort of reason envisaged in the text is instrumental in the sense that you act on an instrumental reason when you return a book in order to keep a promise, or help me out because that is what friendship requires.

have been wrong about myself—that it seemed to me that I felt a trust I turned out not to display—or that I ceased to trust him after I made the report.

Avowing an attitude contrasts with reporting insofar as it sets aside the misleading-mind excuse (pledging an attitude, which ~~also~~ sets aside the changed-mind excuse, will be discussed shortly). Suppose that, in answer to the query, I say: yes, Jones is trustworthy. In that case, by standard conventions, I cannot explain my later acting as if Jones is not trustworthy by saying that I must have got my attitude wrong. Under those conventions, as I understand avowal, I avow my belief that Jones is trustworthy rather than just reporting it. I take a greater risk in communicating my attitude than a report would involve, since I will not be able to invoke the misleading-evidence excuse for a miscommunication. And, of course, I thereby make my communication more credible. Why, you may think, would I opt for an avowal instead of a report unless I was very confident of my trust in Jones?[35]

As I can avow a belief or other attitude in my own name, so can I avow it in the name of others, too. I can assume the authority to speak for them, as in saying 'Jones enjoys the trust of all of us', or something of the kind. If others allow me to avow such an attitude in their name as well as mine, then they will not be able later to say that I got them wrong. They might say this of a journalist reporting on their attitude, but they could not say it of someone they treated as a spokesperson with the right of avowal.

Others would give me the right to make that sort of avowal in their name in the unlikely event that they had made a pledge to act according to whatever attitudes I ascribe to them, at least in a certain domain. In pledging an intention to act in that way, they would each have had to communicate an intention to live up to my words in a way that denies them access ~~to~~ both the misleading-mind and the changed-mind excuse. They could not later excuse a failure to live up to my words either by claiming that they had misread their own intention or by saying that they had changed their minds. In making a pledge, they would have set aside both the misleading-evidence excuse that avowing forecloses and the changed-mind excuse that it still allows.

By Hobbes's account of group agency,[36] individuals who wish to incorporate as an agent—a company, for example, or a commonwealth—must pledge in this way to follow the words of a spokesperson, authorising an individual or a majoritarian body to speak in their name as a group.[37] Outside that special context, however, it is vanishingly unlikely that others might pledge to let me speak for them, allowing me to ascribe the attitude to them on such a basis that they cannot excuse failing to display it by saying that I got them wrong. Why,

[35] For a general, broadly congenial account of avowal, see D Bar-on, *Speaking My Mind: Expression and Self-knowledge* (OUP 2004). For my account of avowing and the pledging mentioned in the next chapter, see P Pettit, 'Making Up Your Mind' (2016) 24 European Journal of Philosophy 3; Pettit (n 7).

[36] T Hobbes, *Leviathan* (E Curley ed, Hackett 1994) ch 16.

[37] P Pettit, *Made with Words: Hobbes on Language, Mind and Politics* (Princeton UP 2008); P Pettit, 'Group Agents Are Not Expressive, Pragmatic or Theoretical Fictions' (2014) 79 Erkenntnis 1641.

independently of the pressures of group agency, would others pledge to act on any attitudes that I ascribe? Why would they authorise me *ex ante* as someone who can speak for their attitudes—even their attitudes in quite a restricted domain—with the status of an avower?

While others are unlikely to give me an *ex ante* authorisation to speak for them, however, there is another way in which I might manage to avow an attitude in their name: I might bet on their *ex post* authorisation and hazard something that counts by conventional or contextual criteria—these will make it clear that I am not merely playing the role of a reporter—as an avowal of a shared attitude. Saying that Jones enjoys all our trust, for example, I might bet on no one objecting to my claiming the role of their spokesperson. And if I get away with the presumption, facing no objection, I will count it as having avowed the attitude successfully in the name of all. Others may not have said 'Yea' *ex ante* to my communication of their attitude but, despite having the opportunity to do so, they will not have said 'Nay' *ex post*. In that sense, they will have virtually, if not actively, authorised my speaking for them.[38]

Given this account of how someone may avow an attitude on behalf of others, it should be clear that any member of Prenormitania will be positioned to avow a favourable attitude on the part of all towards any of the cooperative norms that are perceived and internalised in the community. Given the entrenchment of those norms as a result of their being manifestly perceived and internalised on all sides, it is going to be a safe bet that others will not object to anyone's presuming to avow such an attitude on the part of the society as a whole.

But why would someone be motivated to exercise this capacity for making an avowal in the name of all? Because doing so is bound to be reputationally attractive. The reputational concern, as noted earlier, will give people an incentive to gossip about others who fail to conform, since this will make it manifestly more costly for them to let others down themselves; it will expose them to the charge of hypocrisy. As that motive explains why people will have an incentive to gossip about how far others conform, so it will give them an incentive to avow on behalf of all a favourable attitude towards cooperative norms. By doing so, as by gossiping about others, they will stick their necks out, raise the cost of not conforming themselves and increase the confidence of others that they are reliable conformers.

The members of Prenormitania will not only have the general ability to avow a favourable attitude towards cooperative norms in the name of the society, and they will not only have a personal, reputationally based incentive for doing this: they will also have lots of opportunity for making such avowals. It will be

[38] By standard accounts of conversation, that is what we routinely do as we presume to contribute to a shared body of attitudes in speaking with others, where the absence of objection by any other means that our contribution gets to be registered by each participant as part of that corpus: part of the common ground between us; see R Stalnaker, 'Assertion' in P Cole (ed), *Syntax and Semantics: Pragmatics* (Academic Press 1978); D Lewis, *Philosophical Papers*, vol 1 (OUP 1983) ch 13.

clearly appropriate to do so in teaching the young about how they should behave, in introducing new arrivals to the society and, of course, in reminding offenders of how they have erred. In each case, it will make perfect sense to claim to speak for the society as a whole in claiming that this or that pattern is how we do things around here and how we expect people generally to behave.

Back now to ratification. If any member of Prenormitania has the capacity, the motive and the opportunity to avow in everyone's name a favourable attitude towards cooperative norms, then this is certain to be a matter of common awareness: the evidence that that is so will be accessible to all, accessibly accessible to all, accessibly accessibly accessible to all, and so on. But if it is a matter of common awareness that anyone can avow such an attitude in your name or mine, it will equally be a matter of common awareness that, because we internalise the norm—and not just, for example, because it would be embarrassing to do so—none of us can demur; the conditions that make any such avowal appropriate will make it inappropriate for us to object. As a matter of common awareness, then, we each not only favour cooperative norms individually, but also favour them collectively as members of the community who identify with how things are done there. We ratify the norms together, giving them our shared, social imprimatur.

Ratification in this sense takes the internal point of view further than perception and internalisation, setting up a new dimension of contrast with the perspective of the outsider or outlier. In Prenormitania, it turns out that we not only each see the norms as attractive from an individual perspective, internalising them readily, we also see them as attractive in our social identity, joining with others in supporting and celebrating them. Insofar as the norms have the authority of the community behind them in this way, they will present to us, ironically, as imposing requirements from without. We may favour them in our social identity, but that identity is only part of who we are. So, the very ratification of norms to which we are party in that identity may give them the aspect of obligations to which we are subjected, willy-nilly, in our capacity as individuals.[39]

4. The Personal Commitment that Cooperative Norms Attract

A. The Internal Point of View

The appearance in Prenormitania of cooperative regularities that are generally perceived, internalised and ratified as norms means that the society now well

[39] If Hart's story about primary—and, by extension, secondary—rules explains ratification, then that may help to answer the criticism, roughly, that he cannot account for the fact that when a regime of law is in place, members of the society will take the law to rule on how each ought to behave, at least in important respects, providing a basis for exhorting themselves or others to conformity. Shapiro (n 25) formulates this sort of challenge in one way; for other formulations, see Joseph Raz, 'About Morality and the Nature of Law' (2003) 48 American Journal of Jurisprudence 1; David Plunkett, 'Legal Positivism and the Moral Aim Thesis' (2013) 33 OJLS 563.

and truly deserves the name of Normitania. Those regularities constitute social norms of broadly the kind that Hart casts as primary rules. And they establish the starting point for his genealogy of law: for his argument that Normitania is likely to give way to Lexitania.

The genealogy of Normitania highlights the fact that members of the society are likely to adopt a point of view on their shared norms that is saliently different from that of outside or detached observers; it fleshes out the idea of the internal point of view. Those on the outside will see the norms only as a basis for predicting how members will behave in various circumstances and, if they enter the society, how they will be expected to behave themselves. But the members of Normitania will generally perceive, internalise and ratify them as norms, adopting a stance that marks them off as insiders and, unlike outliers, as more or less willing adherents.

AQ5 There is one further feature of how Normitanians will see those norms that this final section introduces. This feature is bound to appear when conditions like those documented in the two previous sections have been realised; in that sense, it is a derived, not a novel, development. But it is worth highlighting because it reveals a further contrast between their point of view on the norms they sustain and the point of view of outsiders or outliers. It uncovers another dimension in the notion of an internal point of view.

B. Responsibility Ascribed and Invited

The norms considered so far are all socially constructed rules, as Hart emphasises that primary rules are socially constructed. But even though they are social constructs—better perhaps, social precipitates—there is a question as to how far we in Normitania invite others to hold us responsible to them.

The norms are patterns to which we each hold one another to account insofar as we react negatively to defection: we refuse to cooperate with offenders, we gossip about their offences and we indict them with failing to live up to socially ratified expectations. In reacting to defection in such ways, we may be said to hold one another responsible for conformity to the established norms: to hold them to account. And, that being manifest, we each expect to be held responsible for conformity ourselves.

But there is a difference between our each expecting to be held responsible in this sense by others—this may be something we accept but regret—and our each positively inviting others to hold us responsible. To invite others to hold us responsible is to grant them a claim against us—an entitlement to hold us to the norms—recognising that they have an important standing in our lives. And

it turns out that we in Normitania cannot help but issue such an invitation to one another, and confer such an entitlement.[40]

C. Commitment

The paradigm case where I invite responsibility as well as expecting to be ascribed responsibility is marked by the notion of commitment, as it is used in ordinary exchange. When I have committed myself to act in a certain way, I clearly give you a claim over me that I should act in that way. And if I act contrary to my word, then I wrong you by denying you that claim. Suppose I commit myself not to tell a hurtful story about you in certain company and I do then tell that story. I will do you a disservice insofar as I hurt you by the story I tell. But I will do you a distinct disservice insofar as I deny you the claim that my commitment gave you. In virtue of having made the commitment, I owe you my silence on the matter in question, to invoke a familiar commercial metaphor, and I fail to discharge that debt.

The most salient way in which the residents of Normitania might invite others to hold them responsible for conforming to cooperative norms, giving others a distinctive claim against them, would be by committing themselves to live by those norms. In ordinary usage, however, the notion of commitment belongs with morality, being tied up with ideas of right and wrong, obligation and permission. And it would offend against the spirit of Hart's approach to introduce a moralised notion into a genealogical narrative. So is there a non-moral counterpart that might serve the role of such commitment instead? As a matter of fact, there is.

As already noted, to report about one or another of your attitudes—a belief, a desire, an intention or whatever—is to leave open two sorts of explanations for not displaying such an attitude later that may get you off the reputational hook: first, that you were misled about your own mind in making the report; and second, that you changed your mind since making it. Those can each serve as excuses that you might well make if a report about a third person's mind turned out to be false. And if your communications about your own attitudes were of a similar kind—if you spoke about yourself as you might speak about another—then you might claim to explain and excuse a miscommunication on either basis.

[40] For an analogue of the distinction between my expecting to be held responsible to certain norms and my inviting others to hold me responsible, consider the difference marked in recent moral theory between being held responsible from a third-person and a second-person standpoint: S Darwall, *The Second-Person Standpoint: Morality, Respect, and Accountability* (Harvard UP 2006). I may be blamed from a third-person standpoint for doing something that hurts another or that breaches a duty I have, or that offends against a general precept of behaviour. In this sort of case, broadly described, I may be said to do wrong because I do bad. I may be said to do wrong in another more complex case, however, not so much because I do bad—although I may indeed do something that is independently bad—but rather because I act, as we say, in a way that wrongs someone: I act contrary to a claim that by common assent they have over me
.

But, as emphasised, two sorts of speech acts are distinct from attitudinal reports in manifestly setting aside such excuses. The avowal of an attitude sets aside the misleading-mind excuse, as in communicating a belief that Jones is trustworthy by saying simply 'Jones is trustworthy' or an intention to X by saying 'X-ing is the thing to do' or, in a suitable context, 'my intention is to X'. And the pledging of an attitude, which may be possible only with intentions, involves setting aside both excuses at once, as in saying 'Rely on me: I am going to X' or simply 'I promise to X'.[41]

AQ6

Avowals and pledges involve commitment in a game-theory rather than a moral sense. They are acts of communication in which I as speaker choose to stake my reputation on the attitude I self-ascribe being as I characterise it. In the case of avowing an attitude, I foreclose the possibility of excusing a miscommunication by saying: 'I must have got myself wrong: I thought that is what I believed or intended'. In the case of pledging an intention to X—in effect, making a promise to X—I foreclose that possibility and also the possibility of saying 'I changed my mind since speaking with you'.

In each case, it is manifest that I am staking my reputation on being right about the attitude I self-ascribe and, seeing that I am doing this, you are in a position to be much more assured about the presence of the attitude than if I had merely reported it. I make a commitment in the purely descriptive sense of putting my money where my mouth is. I take on a risk that would manifestly lack sense unless I had special knowledge of the attitude: this, presumably, because of having a degree of agential control over whether I develop and display the attitude.[42]

The upshot of these observations is that there is a non-moral sense, perfectly consistent with Hart's genealogy, in which residents of Normitania may commit themselves to abiding by the social norms or primary rules that emerge in their society. They may make a pledge, in the sense introduced, to act as the norms require. And if they do commit themselves in that way, then they may be said to invite one another to hold them responsible to cooperative norms or primary rules.

So is there any ground for thinking that those agents must commit themselves to the emergent norms? Yes, it transpires, there is. But in order to see why that is so, it is necessary to reintroduce an idea also canvassed earlier: that of doing something virtually rather than actively.

[41] Here and elsewhere I set aside complications arising when the content of the avowal or pledge bears on a past time or is highly qualified, say in probabilistic terms. For a fuller treatment, see ~~(Pettit 2018a)~~.

[42] See V McGeer, 'Is "Self-knowledge" an Empirical Problem? Renegotiating the Space of Philosophical Explanation' (1996) 93 Journal of Philosophy 483; R Moran, 'Self-Knowledge: Discovery, Resolution, and Undoing' (1997) 5 European Journal of Philosophy 141; V McGeer, 'The Moral Development of First-person Authority' (2008) 16 European Journal of Philosophy 81.

D. Social Norms as Matters of Virtual Commitment

As discussed earlier, people may authorise you to avow their attitudes in either of two ways: actively or virtually. They will actively authorise you to do this if they say 'Yea' to the proposal that you should speak for them. They will virtually authorise you to do this if you carry on as if you were authorised and if they have the opportunity to object but choose not to do so. In each situation, they exercise control over your speaking for them: in the one case, by saying 'Yea' to your doing so; in the other, by refraining from saying 'Nay' in response to your presuming to do so.

As authorisation and control may be virtual, so too may commitment. Suppose that I am expected to behave in a certain way, as a matter of common awareness between me and those I interact with. And suppose that I do nothing to reject this expectation that is held about me when there is opportunity for me to do so. The fact that the expectation of others is manifest between us—is a matter of common awareness—will mean that I can hardly say, in seeking to excuse the fact that I disappointed them, that they got me wrong or that, while they got me right, I changed my mind after they formed the expectation. Given I knew they hold that expectation, they may reasonably ask why I didn't say anything to remove it. The fact that I said nothing, consciously allowing them to misconstrue my intention, means that I pledged in a virtual sense to behave as they expected me to behave. I didn't say 'Yea' to their expectation, but I could have said 'Nay' and I didn't.

It should be clear that when social norms are fully established in Normitania, being perceived, internalised and ratified as a matter of common awareness, each of us in the society is virtually committed or pledged to complying with them. We are each in a position where, as a matter of common awareness, others are going to expect us to conform unless we distance ourselves from the norms. And if we do not distance ourselves—if we do not say 'Nay' to the expectations—then we must be taken as acquiescing to those expectations and, in effect, committing to live up to the norms.

The upshot of this is that we in Normitania not only sustain the sorts of cooperative norms explored here, perceiving, internalising and ratifying them as norms; we also each virtually pledge ourselves to abide by those norms, making conformity a matter of personal commitment. In other words, we not only expect to be held responsible for conforming to the norms, but also each invite others to hold us responsible; we give others a claim against us, in the event of our failing to conform.

That cooperative norms like those envisaged here are matters of personal commitment in this sense is not an extra feature in the way in which perception, internalisation and ratification count as extra features. The fact that the norms are matters of perception, internalisation and ratification means that they have to display the committal property as well. But it is well worth

emphasising that property, because it highlights yet another respect in which those norms attract an internal point of view on the part of those who sustain them. That you and I commit to the norms, inviting others to rely on our conformity, maximises the contrast between our attitude towards the norms and the attitude of an outsider or an outlier.

5. Conclusion

A. The Upshot

According to the account developed in this article, there is value in extending Herbert Hart's account of how, under suitable conditions, the secondary rules of a legal system would have more or less inevitably emerged from a society governed only by primary rules: how Normitania would have naturally given way to Lexitania. Specifically, there is value in seeking to tell a similar genealogical story as to how primary rules or norms would have emerged themselves: how Prenormitania would have given way to Normitania.

The value in this exploration, so I suggested, is twofold. It vindicates something that Hart presupposes: that there is every reason to think that primary rules or norms would naturally get established in more or less any society. And it helps to give us a sense of how the people among whom they emerge would relate to those rules in the manner of insiders and willing adherents; it substantiates and enriches Hart's claim that people would have an internal point of view on the rules. We saw that, emerging under the impetus of a reputational motor, primary rules would naturally come to be perceived, internalised and ratified by people and that this would mean that they get to be matters of personal commitment on the part of each.

B. Primary Rules and Social Norms

Do the social norms whose emergence we have described really correspond to primary rules, as Hart conceives of those rules? Or do they represent a somewhat different picture of the rules that, by his account, precede the appearance of a properly legal system? The account given of those norms does not aspire to be an exegesis of Hart, as I stressed, not being developed from an interpretation of the things he says about primary rules. But there is still a question as to how far it may call for a revision of Hart's own account of primary rules. The answer to that question, I submit, is that while it may regiment primary rules in some measure, it is not revisionary in any substantive way.

If this is so, then there are two respects in which we should expect to find a fit between social norms, as we have cast them, and primary rules. First, there should be a high degree of convergence between the extension of social norms

and that of primary rules. Secondly, there should be a high degree of convergence between our conception of those norms and Hart's conception of primary rules; his comments on those rules ought to apply in general to norms that are reputationally supported and, to cite their underived features, are perceived, internalised and ratified.

C. The Extensional and Conceptual Fit

The first requirement of fit is that our norms of truth telling, non-violence and non-coercion, keeping promises and respecting property should correspond broadly to what Hart has in mind when he speaks of primary rules. And it is clear that they do, since they impose more or less the same prohibitions or restrictions on people. 'Primary rules,' he says, 'must contain in some form restrictions on the free use of violence, theft, and deception to which human beings are tempted but which they must, in general, repress, if they are to coexist in close proximity to each other.'[43] And he adds immediately, in a phrase redolent of the time, that 'such rules are in fact always found in the primitive societies of which we have knowledge'.

The second requirement of fit is that there is something in Hart's conception of primary rules—indeed, rules in general—that answers to our stipulations that social norms should be reputationally supported and should be perceived, internalised and ratified as norms. He is not explicit on any of these specific issues, but the tenor of the discussion fits very well with the tenor of the presentation here.

Are primary rules reputationally supported for Hart? Well, they are certainly supported, as he argues in various contexts, by 'social criticism and pressure for conformity'.[44] And given that primary rules are 'wholly customary in origin', he says that 'the social pressure may take only the form of a general diffused hostile or critical reaction which may stop short of physical sanctions'. Specifically, the pressure 'may be limited to verbal manifestations of disapproval or of appeals to the individuals' respect for the rule violated; it may depend heavily on the operation of feelings of shame, remorse, and guilt'.[45] The comments come close to acknowledging that reputational motives are crucial in getting people to conform to primary rules.

Does Hart suggest that primary rules are going to be perceived as norms in Prenormitania? Absolutely. With those rules in mind, he implies that they will be taken on all sides as collectively beneficial and individually burdensome, something that could not be the case unless they were perceived as norms by all: registered by all—and, indeed, registered by all as a matter of common awareness. On the one side, he says that they are 'thought important because

[43] Hart (n 1) 90.
[44] Hart (n 1) 57.
[45] Hart (n 1) 86.

they are believed to be necessary to the maintenance of social life or some highly prized feature of it';[46] on the other side, he presumes that it will be 'generally recognized that the conduct required by those rules may, while benefiting others, conflict with what the person who owes the duty may wish to do'.[47]

As he thinks that primary rules are bound to be perceived as norms, so Hart is clear that they are also going to be internalised. When people do not internalise a rule, they abide by it only in the manner of the students in our example of pluralistic ignorance. Hart recognises that where secondary rules have been established, people may differ in the basis on which they conform to any rule, primary or secondary, and that some may even resemble the students in conforming reluctantly, without a preference for general conformity playing any role in their psychology: they may conform to the rules without internalising them. However, he insists that this cannot happen much in a society where primary rules operate in the absence of secondary rules. In such a society, he thinks, 'those who reject the rules except where fear of social pressure induces them to conform' can only be 'a minority'; otherwise they would not be numerous enough to enforce the rules: 'those who reject the rules would have too little social pressure to fear'.[48]

This remark suggests that Hart would take a negative line on the question raised in section 2 concerning whether the internal point of view can enter before internalisation. He would apparently deny that the internal point of view has appeared at the stage where people conform to something they perceive as a norm but only for reputational reasons. Assuming that 'the fear of social pressure' he speaks of in this passage is not just a fear of physical penalty—assuming that it includes a concern with reputation—he would seem to hold that internalisation of a norm is required for the appearance of the internal point of view.

Finally, does Hart think of his primary rules as socially ratified? He does not use that language, of course, but the language that he does use strongly suggests that in our terms those who live under primary rules see them as expressions of a collectively shared attitude within the society. He thinks of the rules as being imposed on people, under the effect of 'social criticism and pressure',[49] 'as a common standard' sponsored from without, not just within.[50] Thus, he holds that people see their requirements as matters of obligation, not as matters of personal interest, inclination or ideal. 'Characteristically, rules so obviously essential as those which restrict the free use of violence are thought

[46] Hart (n 1) 87.
[47] Hart (n 1).
[48] Hart (n 1) 91–2. The insistence that this cannot happen is undermined, of course, by the possibility of pluralistic ignorance, as the earlier discussion makes clear. Hart does not seem to have registered that possibility.
[49] Hart (n 1) 57.
[50] Hart (n 1) 86.

of in terms of obligation. So too rules which require honesty or truth or require the keeping of promises.'[51]

Given the fit between the social norms for which this article has provided a genealogy and the primary rules that Hart posits in his story about law, it is reasonable to claim that what we have done is to extend or elaborate his genealogy. That extension gives a greater temporal reach to the genealogy, as I put it at the beginning, explaining why even a pre-legal society ought to be characterised by something like his primary rules. It also gives a greater psychological depth to the story, providing us with an enriched sense in which it is bound to be true that the members of that society see those rules from within an internal point of view.

[51] Hart (n 1) 87.