Legal Theory, **22** (2016), 1–21.
© Cambridge University Press 2017 0361-6843/17
doi:10.1017/S1352325216000148

NEW DIRECTIONS IN LEGAL EXPRESSIVISM

Matthew X. Etchemendy*
University of Chicago Law School

ABSTRACT
Numerous legal scholars have adopted the view that H. L. A. Hart offered an expressivist (or at least proto-expressivist) account of legal discourse. But philosophers of law have given comparatively little attention to the prospect of developing newer forms of legal expressivism informed by contemporary metaethics. One notable exception is Kevin Toh, who has recently addressed this constructive project. This article critically evaluates Toh's contributions to the project of legal expressivism. Despite the value of Toh's work in reviving legal expressivism as an explanatory strategy in jurisprudence, neither he nor anyone else has paid sufficient attention to the most important question now facing legal expressivists. I call this the "question of content": to wit, what distinguishes the mental states expressed by legal statements from the mental states expressed in other forms of normative discourse? I argue that theorists interested in legal expressivism must prioritize this question.

Legal scholars working in analytic jurisprudence have been increasingly inclined to adopt the view that H. L. A. Hart embraced an expressivist (or at least partly or proto- expressivist) account of legal discourse and thought.[1] Although exegetical questions about the role of expressivism, noncognitivism, and related ideas in Hart's legal philosophy are of some

*I owe a debt of gratitude to Nadeem Hussain, Steve Schaus, Kevin Toh, and Daniel Wodak for their input on drafts of this paper, and to Stefan Sciaraffa for helpful discussion about the themes discussed here. I am also grateful to two anonymous reviewers for helpful comments.

1. Kevin Toh has provided the lengthiest consideration of the role of expressivist and noncognitivist ideas in Hart's work. *See* Kevin Toh, *Hart's Expressivism and His Benthamite Project*, 11 LEGAL THEORY 75 (2005) [hereinafter Toh, *Hart's Expressivism*]. Matthew Adler, Michael Steven Green, Robin Bradley Kar, Stephen Perry, and Scott Shapiro have subsequently adopted the view that Hart was a kind of legal expressivist. *See* SCOTT J. SHAPIRO, LEGALITY (2011), at 98–99; Matthew D. Adler, *Social Facts, Constitutional Interpretation, and the Rule of Recognition*, *in* THE RULE OF RECOGNITION AND THE U.S. CONSTITUTION (Matthew D. Adler & Kenneth Einar Himma eds., 2009), at 193, 200–201; Michael Steven Green, *Kelsen, Quietism, and the Rule of Recognition,*

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

importance because of Hart's overwhelming and enduring influence on modern analytic jurisprudence, there is a strong case to be made that those with an interest in both jurisprudence and expressivism as a general philosophical program have more to gain by attempting to develop *new* forms of legal expressivism[2] that take inspiration from the recent insights of

in THE RULE OF RECOGNITION AND THE U.S. CONSTITUTION (Matthew D. Adler & Kenneth Einar Himma eds., 2009), at 351, 354–361; Michael S. Green, *On Hart's Category Mistake*, 19 LEGAL THEORY 347, 365–369 (2013); Stephen Perry, *Beyond the Distinction Between Positivism and Non-Positivism*, 22 RATIO JURIS 311, 316 n.6 (2009); Stephen Perry, *Where Have All the Powers Gone? Hartian Rules of Recognition, Noncognitivism, and the Constitutional and Jurisprudential Foundations of Law, in* THE RULE OF RECOGNITION AND THE U.S. CONSTITUTION (Matthew D. Adler & Kenneth Einar Himma eds., 2009) [hereinafter Perry, *Powers*], at 295, 309–318; Robin Bradley Kar, *Hart's Response to Exclusive Legal Positivism*, 95 GEO. L.J. 393, 397 & n.17 (2007) (agreeing that "Hart's account of the law *includes* an expressivist dimension," while emphasizing that in his view "it also includes a cognitivist (and hence non-expressivist) dimension"); Scott J. Shapiro, *What Is the Internal Point of View?*, 75 FORDHAM L. REV. 1157, 1169–1170 (2006); *see also* Michael Steven Green, *Legal Realism as Theory of Law*, 46 WM. & MARY L. REV. 1915, 1954 (2005) (noting, in connection with Hart's distinction between internal and external legal statements, that "Hart may . . . have been committed to a form of non-cognitivism about value"). Torben Spaak, in his study of the Scandinavian legal realists, has cautiously endorsed Toh's reading of Hart, concluding that Hart's "sympathies clearly lie with some version of the non-cognitivist theory," TORBEN SPAAK, A CRITICAL APPRAISAL OF KARL OLIVECRONA'S LEGAL PHILOSOPHY (2014), at 11, though Spaak correctly observes that explicit reliance on and invocation of metaethical arguments and premises plays a much smaller role in Hart's work than in Karl Olivecrona's, *id.* at 10–11, 106.

Toh was not strictly the first to notice the expressivist or noncognitivist strands in Hart's work. Joseph Raz was probably the first to interpret Hart this way, at least in print, *see* Joseph Raz, *The Purity of the Pure Theory*, 35 REVUE INTERNATIONALE DE PHILOSOPHIE 441, 448–449 (1981); *see also* Perry, *Powers*, *supra*, at 310 n.53, but Raz's remarks on the subject are comparatively brief. For other early references to the subject of legal expressivism, *see* JULES COLEMAN, THE PRACTICE OF PRINCIPLE (2001), at 88–89 & 89 n.25; Michael Moore, *Legal Reality: A Naturalist Approach to Legal Ontology*, 21 LAW & PHIL. 619, 640 (2002) (describing Hart as "the most famous proponent of a non-cognitivist analysis of legal discourse," but elaborating little on the specifics of Hart's stance). For references to additional early discussions about the possibility of interpreting Hart as an expressivist, *see* Kevin Toh, *Jurisprudential Theories and First-Order Legal Judgments*, 8 PHIL. COMPASS 457, 467 n.11 (2013) [hereinafter Toh, *First-Order Judgments*]. But to the best of my knowledge Toh was the first to address this topic in depth, and he is largely responsible for the recent increase in awareness of, and interest in, the important role of expressivist and noncognitivist ideas in Hart's legal philosophy.

2. Here I use "legal expressivism" to refer to expressivist accounts of legal discourse and thought. The term is sometimes used to refer to theories emphasizing the significance of the so-called expressive dimension or function of laws, or of government action more generally—for example, the possibility that a criminal prohibition might express an attitude of moral condemnation. *See* Matthew D. Adler, *Expressive Theories of Law: A Skeptical Overview*, 148 U. PA. L. REV. 1363, 1364–1375, 1404 (2000); Elizabeth S. Anderson & Richard H. Pildes, *Expressive Theories of Law: A General Restatement*, 148 U. PA. L. REV. 1503, 1504, 1506, 1520, 1531 (2000); Andrew Koppelman, *On the Moral Foundations of Legal Expressivism*, 60 MD. L. REV. 777 (2001). The terminological overlap is basically accidental. *See* Adler, *supra*, at 1380 (noting that such "expressive theories of law" and metaethical expressivism "might incorrectly be understood as related," and explaining the difference between the two).

Toh has suggested in personal correspondence that "meta-legal expressivism" is a preferable term for what I am calling "legal expressivism." (He has also more recently used the term "meta-legal theories" for, inter alia, philosophical accounts of legal discourse and thought. *See* Toh, *First-Order Judgments*, *supra* note 1.) His suggestion is reasonable, given that legal expressivism (as I am using the term) is to legal discourse what *meta*ethical expressivism is to

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

contemporary expressivists like Allan Gibbard, Simon Blackburn, Robert Brandom, and Huw Price. It is time, I think, to move beyond exegesis and to see whether, quite apart from whatever role expressivist ideas may have played in Hart's legal philosophy, expressivism has something to offer us in the search for more illuminating and explanatory accounts of legal discourse and thought. This project—that is, the development of new and improved versions of legal expressivism—is the subject of this paper.

Despite the increasingly wide attribution of expressivist ideas to Hart, only in the last few years has anyone seriously engaged the constructive task of developing new forms of legal expressivism informed by contemporary metaethics. As far as I am aware, in fact, Kevin Toh is the only philosopher who has done so,[3] principally in his article *Legal Judgments as Plural Acceptances of Norms*.[4] I share Toh's interest in the subject of legal expressivism, as well as his belief that a sharper understanding of legal discourse and thought may help clear up some otherwise intractable (and often obscure) jurisprudential debates.[5] But developing a full expressivist account of any sphere of human discourse is no mean feat, as evidenced by the controversies that surround expressivist accounts of moral discourse. We can expect the task to be no easier in the case of legal discourse. Because of these challenges, and because legal expressivism is in its infancy, my goal for the present paper is not to offer a positive proposal for an expressivist account of legal discourse, though I hope to turn to that task in the near future. Instead I will confine myself to a critical assessment of Toh's updated Hartian proposal as developed in *Legal Judgments as Plural Acceptances of Norms*.

The ultimate purpose of this critical assessment is not merely to evaluate Toh's proposal as it currently stands. Rather, I hope to use this paper to make some more general observations about where legal expressivists should focus their efforts given the current state of play. As I will argue, Toh's proposal gets certain fundamental and important things right, but also suffers from a case of misplaced priorities: Toh expends considerable effort attempting to accommodate substantially meritless objections to

ethical/moral discourse and what *meta*normative expressivism is to normative discourse more broadly. Nonetheless, I prefer "legal expressivism" because it is shorter and less cumbersome.

3. Robin Bradley Kar's jurisprudential work also includes expressivist themes, *see* Robin Bradley Kar, *Outcasting, Globalization, and the Emergence of International Law*, 121 YALE L.J. ONLINE 413, 443 n.79 (2012), http://www.yalelawjournal.org/forum/outcasting-globalization-and-the-emergence-of-international-law, but Kar does not appear to straightforwardly conceive of his theory as a form of legal expressivism.

4. Kevin Toh, *Legal Judgments as Plural Acceptances of Norms*, *in* 1 OXFORD STUDIES IN PHILOSOPHY OF LAW, at 107 (Leslie Green & Brian Leiter eds., 2011). Toh's collaboration with David Enoch exploring the possibility that "*legal*" is a "thick" concept could also intersect interestingly with Toh's work on legal expressivism. *See* David Enoch & Kevin Toh, *Legal as a Thick Concept*, *in* PHILOSOPHICAL FOUNDATIONS OF THE NATURE OF LAW (Wil Waluchow & Stefan Sciaraffa eds., 2013), at 257.

5. *See generally* Toh, *First-Order Judgments*, *supra* note 1; Kevin Toh, *Plan-Attitudes, Plan-Contents, and Bootstrapping: Some Thoughts on the Planning Theory of Law*, *in* 3 OXFORD STUDIES IN PHILOSOPHY OF LAW (Leslie Green & Brian Leiter eds.) (forthcoming) (manuscript available at http://ssrn.com/abstract=2307318) [hereinafter Toh, *Plan-Attitudes*].

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

expressivism at nontrivial cost to the philosophical merits of his own proposal, while simultaneously neglecting perhaps the most important task facing legal expressivists, namely that of explaining how the attitudes expressed by legal statements differ from the attitudes expressed by other kinds of statements. Legal expressivists should, I will argue, focus less on trying to rebut general concerns about expressivism (which have, for the most part, been addressed at considerable length in the metaethics literature) and focus more on describing what is *special* about legal discourse and thought, and figuring out how to account for the distinctiveness of legal discourse and thought from within an expressivist explanatory framework.

I. TOH ON PLURAL ACCEPTANCES OF NORMS: AN OVERVIEW

Toh's work on legal expressivism includes an important exegetical component, namely a sophisticated expressivist reading of Hart: Toh has argued for an interpretation according to which Hart's so-called "internal" legal statements are understood to express conative states.[6] But Toh has more recently addressed the more ambitious project of revising Hart's account (as Toh understands it) so as to improve its philosophical merits. Unfortunately space does not permit me to review the reasons why Toh is inclined to think that legal expressivism is worth considering and what kind of philosophical results it can yield. Much of this is set out more than adequately in Toh's earlier exegetical work on Hart, however, and I find many of Toh's points compelling; thus, my present interest is not in reviewing, challenging, defending, or expanding on Toh's rationales for considering legal expressivism philosophically interesting. Rather, my interest is in pointing out the strengths and weaknesses of Toh's recent efforts toward a revised Hartian version of legal expressivism. As noted above, this critical assessment of Toh's recent efforts is meant to serve the broader goal of guiding future work in legal expressivism toward the most pressing philosophical tasks and away from needless distractions.

Toh's revised Hartian proposal employs (and expands on) the idea of "acceptance of norms" as developed by Allan Gibbard in his seminal expressivist monograph *Wise Choices, Apt Feelings*. Toh begins with the following interpretation of Hart's understanding of internal legal statements.[7] According to Toh, Hart holds that such statements:

6. Toh, *Hart's Expressivism*, *supra* note 1, at 77, 79, 88; *see also id.* at 91–105.

7. More strictly, Toh is concerned with *committed* legal statements, which are distinguished from a variant that Raz calls detached legal statements. *See* Toh, *supra* note 4, at 109–111. Toh's own account of detached legal statements (for which *see* Kevin Toh, *Raz on Detachment, Acceptance and Describability*, 27 OXFORD J. LEGAL STUD. 403, 409–414 (2007)), strikes me as basically satisfactory (though it does raise certain subtle questions that cannot adequately be discussed in this paper for lack of space), and I will not be concerned with detached legal statements in the discussion that follows.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

(1) Express the acceptance of a particular norm that is valid according to some other norm R, and
(2) Presuppose that R is accepted and employed as the "fundamental legal norm . . . or rule of recognition" by the officials of the speaker's community.[8]

8. Toh, *supra* note 4, at 116–117. I have rewritten Toh's formulation of Hart's account somewhat. Naturally my goal has been to clarify and not to modify the substance of what Toh has to say. That said, Toh's original formulation presents some exegetical difficulties that leave me feeling less than confident. I will therefore use this note to explain why I believe my reformulation accurately reflects what Toh has in mind.

Toh says that, according to Hart, "in making a [committed internal] legal statement, a speaker (i) expresses his acceptance of a particular norm that is valid according to the most fundamental legal norm of his community, or, as Hart calls it, *the rule of recognition* of his community; and (ii) presupposes that that particular norm is accepted and employed as the rule of recognition by the officials of his community." *Id.*

In constructing my reformulation of the language quoted above, I've interpreted Toh's summary of the first prong as referring to two *separate* norms, viz (1) the "particular norm" whose acceptance the speaker expresses, and (2) the norm the speaker takes to be the "fundamental . . . rule of recognition" (R above). This seems justified for three reasons. First, it is unclear what it would mean for a norm to be valid according to itself, and in any case a rule of recognition that was self-reflexive in this sense would be a strange norm indeed. Second, according to Toh's earlier (2005) formulation of Hart's account, the speaker "(i) expresses his acceptance of R"—where we are told to let R stand for "the norm that [the] speaker considers the rule of recognition of the legal system in his community"—and "(ii) presupposes that R is generally accepted and complied with by the members of his community." Toh, *Hart's Expressivism*, *supra* note 1, at 88. This change in Toh's formulation of the first prong only makes sense if it is motivated by a desire to distinguish the particular norm whose acceptance the speaker expresses from the norm he or she presupposes to be the generally accepted rule of recognition. Third, Toh's more recent formulation of Hart's account in a 2015 paper—though different in certain ways from *both* his 2005 and 2011 formulations—seems to confirm this understanding of the first prong. Kevin Toh, *Erratum to: Four Neglected Prescriptions of Hartian Legal Philosophy*, 34 Law & Phil. 333, 360 (2015) [hereinafter Toh, *Neglected Prescriptions*].

I therefore interpret the referent of the phrase "that particular norm" in the 2011 formulation's second prong as the *second* of the two norms mentioned in the first prong, i.e., the rule of recognition ("R" in my reformulation). This may seem like a strained reading. Arguably a more natural reading would take the phrase "that particular norm" in the second prong to refer to the "particular norm" in the first prong. Still, such an interpretation seems unsustainable for all the reasons mentioned in the preceding paragraph.

The interpretation I've adopted is far more plausible on the philosophical merits than an alternative reading according to which any speaker who makes an internal legal statement expresses the acceptance of a norm that he or she takes to be the fundamental rule of recognition for his or her community. The idea that anyone who makes an internal legal statement expresses his or her acceptance of (and presupposition of the widespread acceptance of) a *fundamental rule of recognition* is very implausible. For such a rule of recognition specifies the feature(s) in virtue of which other rules—e.g., (purported) primary rules of obligation and/or rules of change—are to be deemed legally valid. *See* H. L. A. Hart, The Concept of Law (3d ed. 2012), at 94–96. But most internal legal statements surely do not express acceptance of rules of recognition, but rather acceptance of things like primary rules of obligation. For example, two people might agree that the law requires or forbids a certain action (say, stealing) without agreeing on any rule of recognition; they might accept *different* rules of recognition according to *both* of which stealing is legally forbidden, resulting in agreement concerning the legal validity of the primary rule of obligation in question (that is, the one forbidding stealing) but not the features in virtue of which it is legally valid. Yet surely both individuals would sincerely agree with an internal legal statement to the effect that stealing is illegal. Indeed, this presumably explains why Toh has changed his formulation of the two-pronged Hartian account: Toh's earlier formulation indeed had it that every speaker of an internal legal statement expresses the acceptance of a norm that he or she takes to be the rule of recognition. *See* Toh, *Hart's*

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

Toh goes on to clarify that the factual presupposition component (i.e., the second prong) is "the usual but not an invariable accompaniment of the more primary normative component" (i.e., the first prong).[9] In other words, one who makes a Hartian internal legal statement *normally* expresses a norm-acceptance attitude and presupposes certain natural-factual circumstances. But "even in contexts in which the [factual] condition [normally presupposed] does not obtain and speakers are aware of such presupposition failures, internal legal statements may be used, and to good effect."[10] A good example of this phenomenon—that is, an example where the presuppositional component is absent—might be what I call the case of the deposed king: Consider a deposed king who has been exiled from his country and returns after long decades spent overseas. When he was the recognized king of this country and before he was exiled, he demanded that his vassals display his flag; failures to do so were punishable by a fine. Now he returns to find his vassals (or, at any rate, his former vassals) displaying his usurper's flag instead. Furiously, he exclaims, "My word is law in this land, and I say you are legally obligated to display my flag, not the usurper's." It seems to me that the deposed king could make this statement, in these circumstances, sincerely and without conceptual confusion about the meaning of terms like "law" and "legally obligated." But the deposed king hardly seems to be presupposing that *anyone* in his community accepts a rule of recognition according to which the norms whose acceptance this statement expresses (whatever they may be) are valid (whatever that may mean). In other words he does not presuppose that anyone (though he may *hope* that *someone*) in his community accepts a rule of recognition according to which, as things stand, the members of the community are legally obligated to display his flag. Toh's clarification that the first (normative) prong is the primary or essential component of an internal legal statement and that the second (presuppositional) prong is dispensable thus allows the theory he attributes to Hart to account for cases like this.

Toh attributes this theory to Hart, but he also seems to adopt it as the starting point for the development of his own revised Hartian form of legal expressivism.[11] Toh's major modification of the Hartian account as

Expressivism, *supra* note 1, at 88. Toh's 2011 formulation, as I've interpreted it here, is far more plausible on its merits.

9. Toh, *supra* note 4, at 117.

10. Toh, *Hart's Expressivism*, *supra* note 1, at 113.

11. Toh does not explicitly adopt Hart's account as the starting point for his own, but he does suggest that "[p]rogress can be made," Toh, *supra* note 4, at 118, toward a more philosophically sound expressivist account of legal discourse by combining the Hartian account summarized above with Toh's novel contribution of the notion of "plural norm-acceptances." *See id.* at 118–121. Toh also suggests that Hart, on the above interpretation, offers us "a pretty straightforward expressivist conception of legal statements," albeit one notably hobbled by its (supposed) inability to distinguish "guidings from goadings." *Id.* at 117. To me this suggests that Toh is at least provisionally satisfied with the two-pronged Hartian account—once, of course, it is modified so that the mental states said to be expressed by internal legal statements are plural acceptances of norms rather than norm-acceptance attitudes in the standard Gibbardian sense.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

outlined above is to characterize the conative mental states expressed by internal legal statements not as standard-issue Gibbardian norm-acceptance attitudes but as what Toh calls "attitude[s] of plural norm-acceptance."[12] Toh's characterization of such plural norm-acceptance attitudes is somewhat complicated. Roughly, he says that such plural acceptances of norms involve (inter alia):

(1) An aim on the subject's part to achieve *joint* acceptance of norms (which is to say, normative agreement between himself and his audience).[13]
(2) A "willingness," on the subject's part, "to maintain his normative commitment only if there is, or at least [is] a genuine prospect of, a sufficient uptake on the part of the audience,"[14] i.e., a sufficient prospect of listeners "developing like commitments, partly as the result of [their] recognition of the [subject's] commitment"[15] —at least, that is, if the listeners show a similarly "deferential" or flexible attitude.[16]
(3) A similar (but "*largely* unidirectional") deference to "specialists who have expertise in such normative issues."[17]

I fear this summary may not entirely do justice to Toh's conception of plural acceptance of norms; readers are encouraged to turn to Toh's own description for a better treatment. What is most important to emphasize, however, is that to take an attitude of plural acceptance of norms constitutively involves adopting a particular stance of deference and mutuality toward interlocutors and experts, and a flexible willingness to work toward consensus. Abstracting away from the precise details, this stance of deference, mutuality, and flexibility is the very essence of what distinguishes Toh's plural norm-acceptance attitudes from standard Gibbardian norm-acceptance attitudes.[18] And this is no surprise given that the major

(I discuss the topics of "guiding and goading" and plural acceptances of norms in the portion of the main text immediately following this footnote.)

12. *Id.* at 126.
13. *Id.* at 121–122.
14. *Id.* at 120.
15. *Id.* at 127; *see also id.* at 119–120.
16. *Id.* at 123.
17. *Id.* at 124.
18. Some of Gibbard's remarks in *Wise Choices, Apt Feelings* might leave one with the impression that the effort to develop a novel category of plural norm-acceptance attitudes along these lines is a mistaken enterprise, because Gibbard already thinks that *all* normative statements express psychological states that constitutively involve a certain amount of built-in malleability/persuadability, deference, and/or tendency toward consensus, i.e., that such traits are part of the concept of norm acceptance as developed by Gibbard. After all, Gibbard essentially identifies norm acceptances by their role in a social practice (normative discussion) that, he repeatedly emphasizes, would lose much of its social usefulness (viz, its utility as a mechanism for social coordination) if people were wholly rigid in their attitudes. *See* ALLAN GIBBARD, WISE CHOICES, APT FEELINGS (1990), at 73–75, 77–78 [hereinafter GIBBARD, WISE CHOICES]; *see generally id.* at 64–78. And in *Thinking How to Live*, Gibbard says a "faith . . . that through joint inquiry we can progress from disagreement toward more agreement, and be the better for it . . . is . . . *built into our normative language*." ALLAN GIBBARD, THINKING HOW TO LIVE (2003), at 286 [hereinafter GIBBARD, THINKING] (emphasis added). Does this imply that he sees ordinary norm acceptances as constitutively/necessarily incorporating, at least to some degree, the characteristics Toh stipulatively attributes to plural norm-acceptance attitudes?

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

motivation for Toh's development of this novel twist on Gibbard's norm-expressivism is the "persistent worry . . . that expressivism does a poor job of

I think it would be too hasty to read Gibbard this way. To be sure, natural human psychological dispositions toward mutual influence and consensus play a crucial role in Gibbard's account of the nature and function of normative discourse in *Wise Choices, Apt Feelings*, and a somewhat less prominent role in *Thinking How to Live*. But the question whether normative statements express psychological states that constitutively involve a certain amount of malleability/persuadability, deference, and/or tendency toward consensus is quite distinct from the question whether ordinary, more or less psychologically normal humans *in fact* tend to be malleable/persuadable, deferential (mutually or unidirectionally), and oriented (either willfully or unconsciously) toward consensus when they adopt normative attitudes. One might think that it is an overridingly important and most fortunate fact of the world that most people, most of the time, exhibit some malleability and tendency toward mutual influence in their respective normative attitudes—and even think that normative discourse would cease to serve its core social function were it otherwise—yet *not* think that any degree of malleability or consensus-seeking tendency should be built into the *concept* of accepting a norm.

If nothing else, I am inclined to endorse the latter view on its philosophical merits: I think it yields a more plausible version of comprehensive metanormative expressivism than the alternative. Indeed, it seems to me that to claim that one is not really making a normative judgment unless one has a certain degree of flexibility or consensus-seeking tendency is to trade away a more plausible theory in exchange for one that is less plausible but leaves us with a certain (unwarranted) glow of cheery feeling about the potential for human reconciliation through mutual understanding. But I think it may also be right, as an exegetical matter, to interpret Gibbard as adopting the more plausible (which is to say "thinner") view of what acceptance of norms constitutively involves.

Gibbard appears to see fundamental, rationally irresolvable normative disagreement as a coherent possibility in both *Wise Choices, Apt Feelings*, *see* GIBBARD, WISE CHOICES, *supra*, at 197–199, and *Thinking How to Live*, *see* GIBBARD, THINKING, *supra*, at 268–270, 280–283; *but see* ALLAN GIBBARD, RECONCILING OUR AIMS (Barry Stroud ed., 2008) [hereinafter GIBBARD, RECONCILING], at 174 (expressing hesitation as to whether we "can . . . still think we are disagreeing" when faced with a fundamental impasse). He also seems to think it is possible at least in theory for one to have norm-acceptance attitudes that are immune to persuasive change by means of conversational demands. *See* GIBBARD, WISE CHOICES, *supra*, at 307–308, 309. He thinks we have some control over how flexible we are in our normative attitudes, and how open to persuasion, and that how flexible we ought to be is a substantive normative question, *see id.* at 179–180; GIBBARD, THINKING, *supra*, at 280, so that a *willingness* to be persuaded (one of Toh's criteria for plural acceptance of norms) would not appear to be an essential part of norm acceptance as Gibbard understands it. Moreover, Gibbard's later switch from talk of acceptance of norms to talk of plans, *see id.* at 41–59, 180–181, or at least plan-like attitudes, *see* GIBBARD, RECONCILING, *supra*, at 16, 19—a switch he understands "not as a change of position but as a shift of expository purposes," GIBBARD, THINKING, *supra*, at 181 n.3—suggests that the attitudes expressed by normative statements need not always be flexible or subject to mutual influence. For we have a certain pre-theoretic understanding of what *plans* are, one that Gibbard does not appear interested in disabusing us of: they are not novel theoretical posits like "norm-acceptance" attitudes. (At least not in any respect that seems relevant for present purposes; *see* Allan Gibbard, *Reply to Critics*, 72 PHIL. & PHENOMENOLOGICAL RES. 729, 731 (2006) (identifying certain differences between "planning in the ordinary sense" and "planning" as he uses the term).) And I see no reason to think that people who *plan* must, as such, exhibit flexibility and mutual deference of any sort. If this is right, then Gibbard's repeated emphasis on our persuadability and natural tendency toward consensus may be better understood as anthropological observations about the social function of normative discourse: as noted above, normative discourse would appear to lose much of its social usefulness if people were wholly rigid in their attitudes. But that, of course, is by no means to say that *everyone* who accepts some norms must have some flexibility about his or her stance in this regard. So it seems to me that to the extent Toh's decision to introduce a novel concept of plural acceptance of norms presupposes a certain reading of Gibbard, it is basically a correct reading, or if not, then at least a plausible one.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

making the necessary distinction"[19] between "profferings of reasons on the one hand and instances of browbeating, cajoling, and other such coercive and quasi-coercive proddings on the other"[20]—in short, between "guiding" and "goading."[21] The worry is that legal expressivism may leave us unable to characterize at least some legal disputes—e.g., Dworkinian theoretical disagreements—as anything but shouting matches or "interactions consisting of . . . quasi-coercive influence-mongering."[22]

The worry at issue is not merely that expressivism supposedly leads to a gloomy perspective on normative discourse. The worry is that if legal expressivism implies that Dworkinian theoretical disagreements really are just "quasi-coercive influence-mongering," it would fail, in an important sense, to vindicate a significant part of legal discursive practice.[23] As Toh puts it, we would be unable "to portray discussants involved in [such] dispute[s] in such a way that they would be able to proceed with their dispute openly and in good faith even if they became fully aware of what they are doing in making their respective normative statements."[24] This would be to adopt some kind of error theory toward such disputes, and the burden of proof would be on us to explain how this part of legal discourse could differ so starkly from the participants' own current understanding of it.[25] The introduction of plural norm-acceptance attitudes into Toh's theory is designed to avoid this consequence. (Toh also suggests a second advantage of introducing plural norm-acceptance attitudes into the theory, namely that it helps account for the thought that "commitments to laws are often . . . predicated or dependent on others' having like commitments," while avoiding the problematic aspects of conventionalist theories designed in part to capture this intuition.[26] But this idea is raised in passing and is left undeveloped.)

To summarize, I understand Toh to be offering the following revised Hartian characterization of internal legal statements. Such statements:

(1) Express the *plural acceptance* of a particular norm that is valid according to some other norm R, and
(2) Presuppose that R is accepted and employed as the fundamental legal norm or rule of recognition by the officials of the speaker's community.

In addition to this account of internal legal statements, Toh provides some brief remarks on how he understands Hartian *external* legal statements.

19. Toh, *supra* note 4, at 112.
20. *Id.* at 111.
21. *Id.*
22. *Id.* at 126.
23. Here I use the term "vindicate" as a shorthand way of tracking Toh's concern to portray discussants in a way that they could accept without being unable to continue engaging in the relevant discursive practice openly and in good faith. *See id.* at 111. This differs from, e.g., David Wiggins's use of the term "vindicatory explanation" in the context of explaining moral beliefs. *See* David Wiggins, *Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs*, 91 Proc. Aristotelian Soc'y 61, 66 (1990).
24. Toh, *supra* note 4, at 111.
25. *Id.*
26. *Id.* at 126–127.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

According to Toh, these statements "are in large part *attributions* of norm-acceptances and of their expressions."[27] Although Toh does not elaborate greatly on this, the basic idea seems to be that an external legal statement is a descriptive generalization about the "internal" legal judgments of the members of some relevant reference population.[28] (Note that this could easily be recast in explicitly expressivist terms by saying that external legal statements express ordinarily descriptive beliefs about psychological facts.)

Although Toh's account of external legal statements is not terribly precise, this may be a feature rather than a bug, for external legal statements may themselves involve a substantial degree of vagueness. Additionally, Toh's account of external legal statements seems to do a good job of capturing a familiar and important linguistic phenomenon that metaethicists sometimes describe as the "inverted-commas"[29] use of (a) predicate(s). As Charles Stevenson observed, an anthropologist might say (without any intent to mislead others about his or her attitudes) that "infanticide was morally permissible in Sparta" despite not actually condoning any kind of infanticide, no matter what the cultural or material circumstances.[30] Understood in context, the anthropologist's assertion expresses a belief about psychological facts—something along the lines of the belief that Spartans generally thought that infanticide was morally permissible in certain circumstances. This linguistic phenomenon is very real, far from uncommon, and not at all limited to the case of moral predicates.[31] Many legal philosophers have observed that this phenomenon occurs in legal discourse, but a failure to recognize its considerable significance has arguably led to a great deal of waste and confusion in legal philosophy.[32]

Thus, Toh uses the Hartian term "external legal statements" to refer to statements that use legal predicates in this "inverted-commas" sense.[33] For example, a person might sincerely state that racial segregation was

27. *Id.* at 110 n.5.

28. *See* Toh, *Neglected Prescriptions*, *supra* note 8, at 350 (stating that for Hart, "[e]xternal legal statements are analyzable as descriptions of people's acceptances of the rules that make up their legal system and their behavior motivated by such acceptances").

29. *See, e.g.*, KATARZYNA DE LAZARI-RADEK & PETER SINGER, THE POINT OF VIEW OF THE UNIVERSE (2014), at 301.

30. CHARLES L. STEVENSON, ETHICS AND LANGUAGE (1944), at 83. I have altered Stevenson's example very slightly.

31. For actual examples from the anthropological literature, *see, e.g.*, NAPOLEON A. CHAGNON, NOBLE SAVAGES (2013), at 52 ("But I . . . determined by listening to Yanomamö conversations that you *should not* use a child's name if he or she is sick—it might attract the attention of malevolent spirits . . . ." (emphasis added)); LAWRENCE H. KEELEY, WAR BEFORE CIVILIZATION (1996), at 102–103 (referring to practices such as mutilating enemy corpses and keeping severed enemy ears for trophies as "*no longer* morally . . . acceptable" (emphasis added)); ELMAN R. SERVICE, ORIGINS OF THE STATE AND CIVILIZATION (1975), at 120 ("The Bairu had *no rights* to retaliate by means of blood revenge for a wrong done to them by Bahima . . . ." (emphasis added)); Ernest S. Burch, Jr. & Thomas C. Correll, *Alliance and Conflict: Inter-Regional Relations in North Alaska*, *in* ALLIANCE IN ESKIMO SOCIETY 17, 34 (Lee Guemple ed., 1972) (describing survivors of violent conflicts between Eskimo groups as "morally obliged to seek revenge").

32. *See* Toh, *Plan-Attitudes*, *supra* note 5 (manuscript at 9–15).

33. Toh, *supra* note 4, at 109.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

legally permissible in the United States for many years after the Fourteenth Amendment was ratified, even if the speaker is a constitutional originalist, believes that the Fourteenth Amendment *always* forbade racial segregation, and believes that the Constitution is the supreme law in the United States. Here the statement is best understood as the expression of a belief *about* the attitudes of (some relevant set of) Americans in the past—particularly, the attitudes these people would express by stating (in the non-"inverted commas" sense) that racial segregation was legally permissible in the United States.

This is an important strength of Toh's proposal, and one that is easy to overlook or underappreciate. Although external legal statements play a small part in Toh's discussion—owing, of course, to their "parasitic" or derivative status[34] —their role in the overall theory is crucial. As Toh has persuasively argued elsewhere, a failure to attend to the existence of the "inverted-commas" sense of legal predicates captured by his account of external legal statements can lead to outsize philosophical confusion.[35] Although I see no reason why legal expressivists should stick to Hart's "internal/external" *terminology*—if anything, it might be preferable to jettison it in order to avoid getting ourselves involved in controversies about the proper interpretation of Hart—the underlying *substance* of the phenomenon picked out by Toh's employment of the internal/external distinction is, in my view, quite important.

With this summary in mind, we can now turn to some critiques of Toh's proposal. I will not focus on the gritty details or attempt to critique the proposal in comprehensive fashion. Rather, I will address two major structural issues with Toh's theory as it currently stands. The first issue is the introduction of the idea of plural acceptance of norms, which is motivated by a meritless (or at least apparently meritless) objection to expressivism and impairs the theory's ability to account for essential cases, thereby rendering it vulnerable to obvious counterexamples. The second (and more important) issue with Toh's theory is a crucial *omission*, namely the lack of an adequate account of the content of the attitudes expressed by internal legal statements. I focus on these two structural issues because of the broader lesson they illustrate about where legal expressivists should be focusing their attention given the current state of the field.

II. PLURAL ACCEPTANCE OF NORMS: A SOLUTION TO AN ILLUSORY CONCERN?

The first of the two problems I want to discuss is the introduction of the notion of plural acceptance of norms, the central feature of Toh's most recent work on legal expressivism. Certainly the category of mental states

34. *Id.* at 110.
35. *See* Toh, *Plan-Attitudes*, *supra* note 5 (manuscript at 9–15).

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

Toh demarcates with this label seems common enough, and it may be a sufficiently important species of norm-acceptance attitude to deserve a special designation setting it apart from the broader genus encompassing all norm-acceptance attitudes. But Toh's decision to characterize internal legal statements as expressions of plural acceptances of norms is (1) motivated by an apparently meritless objection to expressivism and (2) renders the theory vulnerable to serious counterexamples.

As discussed above, the impetus for Toh's decision to build plural norm-acceptance attitudes into his theory is the so-called "guiding and goading" problem.[36] The worry is that ordinary expressivist theories, such as standard Gibbardian norm-expressivism, fail to distinguish sufficiently between "profferings of reasons" and "instances of browbeating, cajoling, and other such coercive and quasi-coercive proddings."[37] This is, as Toh says, a persistent worry among certain critics of expressivism. But sophisticated expressivism—and perhaps most pertinently for our purposes, Gibbardian norm-expressivism—has no more trouble than traditional (nonexpressivist) nonnaturalist realism in distinguishing rational discourse from browbeating, to say nothing of coercion. Moreover, the possibility of disagreement so fundamental as to prevent productive rational discourse is a psychological fact having nothing to do with the choice between expressivism and its rivals. Let me turn to a defense of these claims.

Expressivist accounts of moral discourse (particularly noncognitivist ones) have long been met with accusations that they represent a kind of irrationalism. As far as I am aware, though, such accusations rest uniformly on worries about moral *truth*: if there were no moral truths, the thought goes, then moral discourse could not be rational.[38] But why think that expressivism—at least Gibbardian norm-expressivism—leads to moral nihilism, let alone comprehensive normative nihilism? At least at first blush,

36. The name is due to W. D. Falk. *See* W. D. Falk, *Goading and Guiding*, 62 MIND 145 (1953).

37. Toh, *supra* note 4, at 111. Though it is a somewhat tangential point, it is worth noting that Toh's formulation of the problem—which, I think, pretty accurately captures the thinking of those drawn to "guiding versus goading" worries—assumes a false dichotomy between rational discourse on the one hand and browbeating, cajoling, and coercion on the other. Plenty of nonrational persuasive interactions are perfectly civil, nonthreatening, and free of any unctuousness or wheedling. Certain poems of Ralph Waldo Emerson might, for example, instill in readers a greater passion for enjoying and preserving nature. This would presumably not be an instance of *rational* persuasion—or at least in some cases would not be—yet few readers of transcendental poetry feel browbeaten or cajoled, let alone coerced.

38. Strictly speaking this is false, since if "rational discourse" means discourse that consists of the sincere proffering of reasons, then moral discourse could still be rational even if there were no moral truths. There can, after all, be good reasons to have false beliefs or to acquire (or retain) non-truth-apt attitudes. That said, if noncognitivist expressivism led to moral nihilism, it would be hard for it not to slide into comprehensive normative nihilism, for noncognitivist versions of metanormative expressivism generally characterize *all* normative language in expressivist terms and *all* normative thought in noncognitive terms. And it certainly would be impossible to sustain rational moral discourse if there were no *normative truths at all*, for then there would be no genuine reasons to be proffered; rationality itself would be an illusion. *See* Nadeem J. Z. Hussain, *Error Theory and Fictionalism*, in THE ROUTLEDGE COMPANION TO ETHICS (John Skorupski ed., 2010), at 335, 340–341.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

all that is required to avoid such a consequence is to adopt a deflationist theory of truth, or something akin to it.[39] Once we make this basic deflationist move, there is no mystery about how expressivism can vindicate not just moral talk generally but also talk of moral truth.[40] Roughly speaking, to assert that a moral proposition is true is equivalent to asserting the moral claim itself—perhaps with some added normative "oomph."[41] If something along these lines is right, there is no difficulty reconciling expressivism with a commitment to moral truth, since there is no difficulty reconciling expressivism with the underlying moral claims whose truth we would affirm.[42] I see no reason at all why deflationism should not be equally effective as an antidote to any similar worries about *legal* expressivism.

Once we take this deflationist route to moral truth, the expressivist's answer to the question of how to distinguish rational discourse from nonrational discourse is simple enough. If you want to debate things rationally, you should try to convince your audience by offering them genuine normative reasons to agree with you. If not (and bracketing for a moment the moral permissibility of nonrational persuasive influence in various contexts and by various different methods), your repertoire may include flattery, appeals to wishful thinking, threats of physical violence, or whatever else gets you the results you want.

As for the more specific problem with which Toh is concerned—normative disagreements between individuals committed to conflicting or contradictory *fundamental* norms[43] —the same answer applies. The distinction between guiding and goading is that guiding consists of offering genuine normative reasons (or at least sincerely trying to do so). The only evident difference in the context of such fundamental disputes, and the difference that makes such disputes so intuitively troubling, is that neither party is likely to recognize the other party's proffered reasons *as* genuine reasons. In other words, it's easy enough to *guide* one's interlocutors (in Toh's sense) even in the case of fundamental disputes, but there is an acute risk that they will not *follow*—or that if they do, they will do so only out of fatigue or a desire to conform, rather than because they perceive the rational weight of the arguments on offer. This may be an unpleasant thought, but it is apparently a genuine psychological possibility. The real question is this: Is there anything we would *want* to say about the prospects for a rational resolution to fundamental normative disputes that (1) expressivists cannot

39. For an interesting account of truth that would work well here, *see* HUW PRICE, NATURALISM WITHOUT MIRRORS (2011), at 163–183.

40. Of course an expressivist might accept deflationism about truth and *still* disagree that there are any moral truths, but such a combination of expressivism and moral nihilism would be accidental insofar as the former in no way commits one to the latter.

41. *See* PRICE, *supra* note 39, at 163–183.

42. *See* SIMON BLACKBURN, RULING PASSIONS (1998), at 75–83; MARK SCHROEDER, NONCOGNITIVISM IN ETHICS (2010), at 151–165; Paul Horwich, *The Essence of Expressivism*, 54 ANALYSIS 19, 19–20 (1994); Paul Horwich, *Gibbard's Theory of Norms*, 22 PHIL. & PUB. AFF. 67, 70–74 (1993).

43. Toh, *supra* note 4, at 112.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

say, but (2) at least some opponents of expressivism *can* plausibly say? As far as I can see, the answer is no.[44]

44. In discussing this issue, I have interpreted the term "guiding" as synonymous with "proffering genuine normative reasons." This is the interpretation that the text of Toh's paper strongly suggests. *See* Toh, *supra* note 4, at 107–108, 111. So understood, the "guiding and goading" problem is relatively narrow and manageable. It is important, lest my expressivist response to this problem seem too quick, to recognize that the narrow issue involved is distinct from a variety of other, at least thematically related, objections or problems.

First, as noted in the main text, the expressivist reply I have sketched would involve acknowledging that guiding might well be inefficacious in certain "fundamental" disputes. But this in no way diminishes the expressivist's ability to distinguish guiding—the proffering of genuine normative reasons—from other forms of persuasion or psychological influence. (It bears emphasizing here that *inefficacy* is not the same as *impossibility*: an expressivist need not be any kind of existence internalist about reasons, so there is no tension involved in suggesting that fundamental disputes may often involve the proffering of genuine normative reasons that, despite their status *as* genuine normative reasons, simply fall on deaf ears.) And while the conclusion that guiding might well be inefficacious in certain fundamental disputes does, in a sense, "portray discussants involved in [some] moral or legal dispute[s] in such a way that they would" not "be able to proceed . . . openly and in good faith even if they became fully aware of what they are doing," *id.* at 111, it does not represent an error theory in the traditional sense: the claim is not that the statements discussants make—even in the most intractable and fundamental moral or legal disputes—are systematically false or nonsensical. The claim is only that such interactions may be unproductive in some more diffuse, pragmatic sense. So while it might be unpleasant to contemplate the possibility of cases in which efforts to "guide" must simply fail to get off the ground, it would take much more work to develop, on this basis, a convincing philosophical worry for expressivism (let alone for expressivism *in particular*).

It is also important to recognize that questions about the possibility and efficacy of guiding (which is to say, the possibility and efficacy of proffering genuine normative reasons) are distinct from the following, much deeper issue: whether expressivists can provide a convincing account of the *phenomenon of disagreement itself*—i.e., whether expressivists can adequately explain what differentiates disagreements from other differences in attitude, either in the context of fundamental disputes or more generally. (For a recent discussion of this issue, *see* Mike Ridge, *Disagreement*, 86 PHIL. & PHENOMENOLOGICAL RES. 41 (2013); *see also* Michael Bratman, *Normative Thinking and Planning, Individual and Shared: Reflections on Allan Gibbard's Tanner Lectures*, *in* GIBBARD, RECONCILING, *supra* note 18, at 91, 98–100.) The two issues are arguably related in a (very broad) sense, and at times Toh contrasts "genuine normative disagreements" with "sessions of mutual browbeating," Toh, *supra* note 4, at 115, or "instances of talking at cross-purposes," *id.* at 126, which suggests a concern with the latter, deeper issue. Nonetheless, the "guiding vs. goading" problem is at least analytically distinct from what we might call the "disagreement vs. mere difference" problem. That is to say, the problem of drawing a sufficiently clear distinction between guiding and goading is not equivalent to the problem of drawing a sufficiently clear distinction between disagreement and other differences in attitude. Consider, by way of illustration, two interlocutors, A and B, who are discussing the morality of capital punishment. The two interlocutors may *disagree* about whether capital punishment is moral; e.g., where Interlocutor A thinks capital punishment is moral and Interlocutor B thinks it is immoral. Alternatively, they may only *differ* in their attitudes toward the morality of capital punishment, but not *disagree*; e.g., where Interlocutor A thinks capital punishment is moral and Interlocutor B simply has no opinion about the moral status of capital punishment. Note, however, that *in either case* A could, in an effort to get B to come around to A's way of thinking, either guide B (e.g., by saying, "the prospect of capital punishment surely deters numerous would-be killers") or goad B (e.g., by just repeating "capital punishment is good" in a loud voice).

Providing an adequate account of the distinction between disagreements and other differences is a hard and complex problem, one that I cannot address here. But it is not the problem Toh explicitly sets out to address, nor is it clear whether, and if so how, the introduction of plural acceptance of norms into the theory would help on this front.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

All of this, I think, is a fairly standard contemporary expressivist riposte to familiar worries of the "guiding and goading" genre. I also happen to think that this deflationist line of argument is satisfactory and that "guiding and goading" objections are meritless. True, I cannot rule out the possibility that this standard response is subtly inadequate. Perhaps it will emerge that contemporary expressivists have been too glib in their deflationary dismissals of such concerns. Be that as it may, the introduction of plural norm-acceptance attitudes into the theory comes with immediate and readily apparent theoretical costs. Toh recognizes that building the idea of plural acceptances of norms into his theory leaves (or at least may plausibly appear to leave) his account "obviously vulnerable to counter-examples." As he explains:

> There may be, and actually are, people, and even judges, whose legal judgments would be unaltered by their community's clear and stable adherence to contrary legal positions. We can imagine a judge whose opinion that abortion is illegal would persist even in the face of mounting judicial and societal consensus for the legality of abortion, and no real prospect of either ever changing. To think that such a judge is mistaken is one thing; but my conception of legal judgments has the seemingly implausible implication that such a judge is not even making a legal judgment.[45]

Precisely so. Still, Toh bites the bullet, giving voice to his intuition that such a judge would *not* actually be making a legal judgment.[46] To my mind, this is an astonishing conclusion, one that certainly finds little if any support from actual usage. (Of course, here I must simply depend on readers' intuitions agreeing with my own, but I suspect that they will.)

Toh himself seems uncomfortable with this conclusion about the recalcitrant judge. He confesses that his "intuition may be corrupted by [his] own theory," and hastens to add that the goal of his account is not to get things perfectly counterexample-proof, but just to do a reasonably good job of capturing the central features of the discursive practice.[47] Toh's realism about how well we can expect a theory of legal discourse to fit the data, viz, rather less than perfectly, is well taken. There is nothing wrong with adopting a theory that requires one to acknowledge a nontrivial counterexample or two. But given the apparent ease with which sophisticated expressivists can turn aside "guiding versus goading" worries, introducing the notion of plural acceptance of norms into the theory (with all the attendant costs to the plausibility of the theory) seems unjustified. It would seem preferable simply to proceed along more traditional Gibbardian lines unless and until "guiding versus goading" worries prove more substantive than they presently appear.

45. Toh, *supra* note 4, at 133 (footnote omitted).
46. *Id.*
47. *Id.* at 133–134.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

III. THE MISSING ELEMENT: ADDRESSING THE QUESTION OF CONTENT

This brings us to the second problem with Toh's theory as it now stands. Whether the theory characterizes internal legal statements as (perhaps inter alia) expressions of standard Gibbardian norm-acceptance attitudes or plural norm-acceptance attitudes instead, an overwhelmingly important question remains. What norms, precisely, do internal legal statements express the acceptance of? We can call this the "question of content," and it is almost certainly the most pressing question for any theorist who wishes to build an account of legal discourse on a Gibbardian norm-expressivist foundation. This is true whether we are characterizing internal legal statements as expressions of standard Gibbardian norm-acceptance attitudes or as expressions of some more exotic flavor of norm-acceptance attitude, like plural acceptance of norms.

Somewhat surprisingly, Toh does little to address this core question. Recall that according to his account, the real sine qua non of an internal legal statement is that it expresses the (plural) acceptance of a particular norm (call it N) that is valid according to some other norm R. Because every norm is presumably valid according to *some* other norm (whatever "valid" here means), all this really tells us is that an internal legal statement expresses the (plural) acceptance of a norm N. But this leaves Toh's theory strikingly unexplanatory. To start with what I take to be an obvious point, the statement that "the thing to do is not to steal" and the statement that "you are legally obligated not to steal" (or that "the law forbids you from stealing") express different attitudes. So if we characterize both of these statements as expressions of the acceptance of norms—as we will presumably do if we are building on a broadly Gibbardian norm-expressivist foundation—the only apparent way to distinguish them is on the basis of the *content* of the norms whose acceptance they express.

By analogy, a statement that "the thing to do is not to steal" and a statement that "it is immoral to steal" seem to express different attitudes, at least if morality is not conceived so broadly as to encompass all of practical reasoning. So again, if we characterize both of these statements as expressions of the acceptance of norms, the only apparent way to distinguish them is on the basis of the content of the norms whose acceptance they express. For his part, Gibbard accomplishes this by identifying a class of distinctly moral norms governing certain emotional responses. Hence the statement that "no one should ever steal" expresses the acceptance of norms governing the act of stealing; the statement that "stealing is always immoral," by contrast, expresses the acceptance of norms governing feelings of guilt on the part of thieves, and feelings of indignation or anger toward thieves.[48]

48. *See* GIBBARD, WISE CHOICES, *supra* note 18, at 40–45; *see also* GIBBARD, RECONCILING, *supra* note 18, at 16–17 (reiterating and refining this account of moral statements).

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

My point is not to endorse the details of Gibbard's account, but only to offer an example of the *kind* of story we'd have to tell in order to provide a satisfactory norm-expressivist account of internal legal statements, which would be analogous to the story that Gibbard tells in *Wise Choices, Apt Feelings* in order to provide a norm-expressivist account of moral statements.

An adequate version of legal expressivism must be able to distinguish the attitudes expressed by the statement that "no one should ever steal," the statement that "stealing is immoral," *and* the statement that "stealing is unlawful," for the simple reason that these statements rather plainly do not express the same attitude.[49] Now it may *in fact* be that the requirements of morality and of law are identical, or that one is a subset of the other: such notions are philosophically implausible but not conceptually incoherent. But at the same time, it is clear that such an identity is not a conceptual truth. It just is not plausible to claim that a common law appellate judge who opines, "Our precedents make it clear that, however immoral it may be, lying is not, in the ordinary case, unlawful," is speaking self-contradictory nonsense. Nor is a judge who says, "Though parking in a handicapped zone at 3 a.m. in an empty lot is not immoral, it is still illegal, and so it is proper for me to fine a person who does this." Thus, we must be able to explain the difference between the content of those norms whose acceptance is expressed by the claim that, for some X, "X is immoral," and the content of those norms whose acceptance is expressed by the claim that, for the same X, "X is unlawful."

Toh is aware that a plausible form of legal expressivism must be able to explain these differences, and of the general strategy by which to do so from within a broadly Gibbardian framework that characterizes basic or "all-things-considered" normative claims (e.g., "X is the thing to do"), moral claims, and legal claims alike as expressions of the acceptance of norms.[50] That is to say, he agrees that the correct strategy is to distinguish those norms whose acceptance is expressed by legal statements from those whose acceptance is expressed by moral statements and other kinds of normative statements. (Indeed, Toh specifically mentions Gibbard's account of moral statements in this context.) Yet Toh makes only a brief nod toward precisely *what* the difference between these norms is. He says,

> Legal statements . . . probably do not deal with norms governing guilt and impartial anger, or any other kind of emotions. Instead, according to Hart,

49. Robin Bradley Kar, whose jurisprudential work includes expressivist themes, *see supra* note 3, has shown that he too is aware of the need to distinguish the attitudes expressed by these different kinds of statements. His own comments on what I've called the question of content are, however, brief and schematic. (They also appear to bear a marked thematic resemblance to Toh's similarly brief take on the subject.) *See* Robin Bradley Kar, *The Deep Structure of Law and Morality*, 84 TEX. L. REV. 877, 940–941 (2006); Robin Bradley Kar, *The Two Faces of Morality: How Evolutionary Theory Can Both Vindicate and Debunk Morality (with a Special Nod to the Growing Importance of Law)*, *in* EVOLUTION AND MORALITY: NOMOS LII (James E. Fleming & Sanford Levinson eds., 2012), at 31, 89–90; *see also* Kar, *supra* note 3, at 459–460.

50. Toh, *supra* note 4, at 130–131.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

> they deal with norms of the following kind: modifications of these norms, resolutions of disputes about contents of these norms, and determinations about the applicability of such norms are all amenable to being carried out authoritatively by a set of institutions. These are not characteristics of moral norms.[51]

But these brief remarks are both puzzling and quite unsatisfactory. Even if a legal statement expresses not just the acceptance of a particular norm N but also the acceptance of certain norms about how N is to be "modified" or how disputes about N's applicability are to be resolved, we should like to know what the content of N itself is.[52] Indeed, we *need* to know the answer to this question in order to capture adequately the difference between legal statements and other kinds of statements (e.g., moral statements) that express acceptance of norms.

To see why, let us consider a thought experiment. Suppose two interlocutors, call them Discussant A and Discussant B, are discussing the subject of recreational marijuana use. Let us suppose that they agree on all of what we might loosely term the "plain facts" pertinent to the subject, but disagree about whether recreational marijuana use is legal. Discussant A believes that recreational marijuana use is unlawful, and Discussant B believes that recreational marijuana use is lawful. (Perhaps, for example, both discussants agree that the legislature at least purported to pass a law forbidding recreational marijuana use, but Discussant B, unlike Discussant A, believes that a statute forbidding recreational marijuana use would be unconstitutional and hence void.) In other words, Discussants A and B accept, respectively, the norms whose acceptance would be expressed by internal legal statements to the effect that "recreational marijuana use is unlawful" (Discussant A) and "recreational marijuana use is lawful" (Discussant B).

Just what are these norms? Well, we cannot conclude, based solely on what we know so far about Discussant A and Discussant B, that Discussant A accepts norms prohibiting recreational marijuana use. Otherwise the judgment that recreational marijuana use is legally forbidden (right now) but that using marijuana recreationally is still the thing to do (right now) would be incoherent. But however blameworthy the judgment that "I am legally obligated not to use marijuana recreationally, but it still makes sense for me to smoke the marijuana cigarette now in my hand" might (or might

51. *Id.* at 131 (citation omitted).

52. There are further problems with Toh's brief remarks on the question of content. Norms cannot be modified; they are, if I may indulge a bit of metaphysical jargon, abstract entities. Perhaps the thought is not that institutions can *modify* the norms in question, but that these norms forbid or require actions only when certain conditions related to institutional procedures obtain. But even so, these brief remarks raise far more questions than they answer. For example, in order to believe that doing X is unlawful, must one believe that some institutional procedure can *change* whether doing X is lawful? This is arguably nothing more than a modern prejudice—a perfectly coherent position to take, of course (and quite possibly the correct one), but not one built into the very idea of a law or of a legal obligation.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

not) be, it certainly does not seem incoherent[53] —certainly no more so than the judgment that "doing X is immoral, but it still makes sense for me to do X forthwith."[54] It would be quite problematic, therefore, if our theory told us that Discussant A must accept norms forbidding recreational marijuana use. And note that it would not help to posit that Discussant A may accept norms forbidding recreational marijuana use only on the condition—satisfied, by Discussant A's lights—that the legislature has passed a valid statute banning recreational marijuana use. For if nothing else, we want to make room for the possibility that Discussant A thinks that using marijuana recreationally is sometimes the thing to do even when the legislature has passed a valid statute banning recreational marijuana use.

Similarly, and perhaps even more obviously, we cannot conclude that Discussant B accepts norms permitting recreational marijuana use. After all, even if recreational marijuana use is lawful, using marijuana recreationally may still rarely or never be okay to do. (Consider, for example, that committing adultery, gratuitously insulting loyal friends, sadistically picking apart insects just for fun, committing suicide when your favorite football team loses a game, and many other equally nasty or pointlessly destructive things are perfectly legal in a wide range of circumstances. Surely that does not imply that such actions are *ever* okay to do.)[55] Addressing the content of the norms whose acceptance is expressed by internal legal statements should therefore be at the top of our priority list if we are to develop an expressivist account of such statements that characterizes them as expressions of the acceptance of norms.[56]

Once again, however, I think the most important lesson to draw from this discussion is not about the merits and flaws of Toh's proposal; rather, it is a lesson about where legal expressivists should be focusing their attention. *Any* form of legal expressivism, whether it builds on a Gibbardian norm-expressivist foundation or not, must be principally concerned with explaining what is special about legal discourse and thought. Given the current state of the field—in particular, given the relatively undeveloped state of contemporary legal expressivism—this should be the primary concern of legal expressivists today.

53. *Cf.* David Enoch, *Reason-Giving and the Law*, *in* 1 Oxford Studies in Philosophy of Law (Leslie Green & Brian Leiter eds., 2011), at 1, 22 ("[T]here is no contradiction in saying that the law requires that you Φ, but you don't have any genuine reason to Φ.").

54. Indeed, even if one did think that the latter (moral) claim is incoherent, it would be quite another matter to suggest that the former (legal) claim is also incoherent. It seems to me that the internal tension appears rather less severe in the former (legal) case.

55. My use of the phrase "okay to do" in this context tracks Gibbard's. *See* Gibbard, Reconciling, *supra* note 18, at 19–20; Gibbard, Thinking, *supra* note 18, at 137–138.

56. Of course, these examples are just the tip of the iceberg: a really well-developed form of legal expressivism should probably have *something* to say not just about statements like "stealing is unlawful" but also statements like "you have a legal right to dispose of your own property," and any number of other kinds of internal legal statements. If we want to say that all of these express the acceptance of various different norms, the norms are bound to get pretty complex pretty quickly. This is not meant to be discouraging, but only to show just how much work legal expressivists have left to do.

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

How might legal expressivists get a start on this problem? Unfortunately it is beyond the scope of this paper to say much about this question; my main concern has simply been to point out its existence and its overriding centrality to the legal expressivist project. Nonetheless, a few brief suggestions may be in order. First, I am not convinced that legal expressivists should try to characterize internal legal statements as expressions of Gibbardian norm-acceptance states. Even so, I understand the attraction of this strategy given the richness and prominence of Gibbard's norm-expressivist program. Supposing, therefore, that we *were* to stick with the norm-expressivist strategy, I think Toh has gestured in roughly the right direction with his reference to authoritative institutional action. Indeed, although the suggestion may seem vaguely *outré* given current intellectual trends in analytic jurisprudence, I think one of the more promising ideas for legal norm-expressivists to consider—and one that certainly would draw on a long line of philosophical tradition—is that internal legal statements express the acceptance of norms governing the use of institutional coercive force.

Unfortunately, fuller development of this idea will have to wait. But for present purposes, it is worth pointing out how this institutional-coercive-force-based approach is the *sort* of tactic that could at least satisfy the minimum condition of distinguishing the norms whose acceptance is expressed by internal legal statements, moral statements, and basic or "all-things-considered" normative statements, respectively. Although it may seem (and indeed would be) simplistic to suggest that the statement "stealing is against the law" expresses acceptance of norms permitting the use of institutional coercive force to prevent or punish stealing, it would at least have the virtue of showing how the statement "stealing is against the law" differs from the statement "stealing is immoral": assuming a Gibbardian account of moral discourse, the latter would express the acceptance of norms governing certain emotional responses to thievery, whereas the former would express the acceptance of norms governing certain coercive institutional responses to thievery. It is simple enough to see, therefore, how such a theory could make sense of the fact that one could coherently think that stealing is both immoral and illegal, immoral but not illegal, illegal but not immoral, or neither immoral nor illegal. And it is also simple enough to see how such a theory could explain the coherence of the thought that it is sometimes, perhaps even often, rational to break the law. (For example, a system of norms that permitted a person to steal but also permitted institutional authorities to coercively punish that person for stealing might be strange, but it would certainly not be incoherent.) Again, my point is not to defend this very simple version of legal expressivism—which is, in truth, quite inadequate—but only to use it to illustrate how the question of content might coherently be addressed from within a legal norm-expressivist framework, and to point in one promising direction (which I hope to explore in further detail later).

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press

IV. CONCLUSION

Although I have focused on the merits and flaws of Toh's recent work on legal expressivism, the ultimate goal of this paper is broader and more forward-looking. After all, the variety of possible expressivist accounts of legal discourse is practically infinite. As legal expressivists, we might characterize some legal statements as the expression of conative attitudes, or we might not. We might characterize some legal statements as the expression of the acceptance of norms in a Gibbardian sense, or we might not. Legal expressivism is a wide and almost entirely uncharted philosophical territory. Hart (and other legal philosophers of past decades whose theories were influenced by noncognitivist ideas, like Karl Olivecrona)[57] did not have the benefit of the considerable advances metaethical expressivists have made in recent years. And as far as I know, Toh is the only contemporary philosopher of law who has made progress toward a version of legal expressivism that incorporates the insights of contemporary metaethical expressivism. It is less important at this stage to critique in minute detail the avowedly preliminary[58] efforts Toh has thus far undertaken than it is to get a good handle on the tasks that are most pressing (and also the tasks that are *not* pressing) if we are to make progress toward better, more illuminating forms of legal expressivism. Toh's work on legal expressivism is to be commended for its sophistication and originality, but it also suffers from—and illustrates the problems with—an excessive focus on general worries about expressivism as an *overall* philosophical program (which, as the example of the "guiding vs. goading" worry illustrates, have already received detailed treatment in the metaethics literature). Of course, there is no reason why legal expressivists should not concern themselves with such general objections to expressivism. But efforts to address such concerns should not come at the expense of addressing the practically untouched question of how to capture the distinctive characteristics of legal discourse from within a broadly expressivist explanatory framework.

57. On the role of noncognitivist ideas in Olivecrona's legal philosophy, *see* KARL OLIVECRONA, LAW AS FACT (2d ed. 1971), at 182–185, 215–216, 259–267; KARL OLIVECRONA, LAW AS FACT (1st ed. 1939), 21–22, 95–96; SPAAK, *supra* note 1, at 94–97, 106, 133, 148–150; Torben Spaak, *Karl Olivecrona's Legal Philosophy. A Critical Appraisal*, 24 RATIO JURIS 156, 163–165, 174–177 (2011); Torben Spaak, *Naturalism in Scandinavian and American Realism: Similarities and Differences*, *in* DE LEGE, UPPSALA–MINNESOTA COLLOQUIUM: LAW, CULTURE AND VALUES (Mattias Dahlberg ed., 2009), at 33, 52–55.

58. *See* Toh, *supra* note 4, at 107 (describing the paper *Legal Judgments as Plural Acceptances of Norms* as "a sort of progress report" requiring "further working-out").

https://doi.org/10.1017/S1352325216000148 Published online by Cambridge University Press